IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 March 2024, accepted 27 May 2024, date of publication 5 June 2024, date of current version 10 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3409762

RESEARCH ARTICLE

Artificial Intelligence in Finance: Coffee Commodity Trading Big Data for Informed Decision Making

NGOC-BAO-VAN LE[1,2], YEONG-SEOK SEO[3], (Member, IEEE), AND JUN-HO HUH[2,4], (Member, IEEE)

[1]Department of Data Informatics, National Korea Maritime and Ocean University, Busan 49112, Republic of Korea
[2]Interdisciplinary Major of Ocean Renewable Energy Engineering, National Korea Maritime and Ocean University, Busan 49112, Republic of Korea
[3]School of Computer Science and Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea
[4]Department of Data Science, National Korea Maritime and Ocean University, Busan 49112, Republic of Korea

Corresponding authors: Jun-Ho Huh (72networks@kmou.ac.kr) and Yeong-Seok Seo (ysseo@yu.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (MSIT) under Grant NRF-2023R1A2C1008134.

ABSTRACT Coffee, the second-largest global soft commodity, can take advantage of a comprehensive mining of daily and historical market data for more effective informed trading decisions. Advanced ICT and data mining technologies can change the trading market operation. The existing systems are confronted with certain constraints, including incomplete data, insufficient documentation for storage, and a requirement for a scalable infrastructure for big data analytics, such as a data warehouse or data lakehouse. To address this issue, the paper presents a design and implementation of a coffee commodity trading big data warehouse capable of analyzing various essential parameters for supporting informed decision-making. First, the designed system can automatically collect coffee trading data for New York Arabica coffee futures prices from selected worldwide reports and financial data portals. Next, the Extract, transform, and load (ETL) process is adopted to ingest coffee futures trading crawled data into the 3 layers data warehouse. Finally, the analytical system will extract and visualize selected key dimensions that influence coffee futures prices within different observation windows and perspectives. As a result, we implement a prototype of a coffee trading data warehouse on the crawled data from January 2000 to October 2022 and visualize trends in coffee futures prices based on the collected data for informed decision-making. The construction system is capable of stably operating and processing large volumes of transaction data. This paper will be valuable documentation for reference and decision support for coffee commodity trading enterprises and contribute to the development of future forecasting algorithms.

INDEX TERMS Coffee big data, data warehouse, coffee commodity trading, ETL process, informed decision-making, data visualization, big data.

I. INTRODUCTION

Coffee has become the product with a high level of consumption worldwide in recent years, with over 2.25 billion cups of coffee consumed daily [1] and playing an essential role in many countries. For example, in Vietnam, over the years, coffee has been the main export and become the second most exported goods in export turnover after rice. Every year, coffee contributes up to 10% of the country's total export turnover [2]. Trading coffee through an exchange is not a new type of transaction. However, traders and investors will still be exposed to many risks if they do not adequately understand the market and the trading methods, price trends, or market data analysis. Price fluctuations, exchange rates [3],

The associate editor coordinating the review of this manuscript and approving it for publication was Walter Didimo.

and export prices of big exporting countries can affect coffee prices in domestic countries. Price fluctuation in world coffee trading and the changes in the exchange market should be monitored carefully in every session and compared to the trading historical periods. Coffee trading exchange data is presented as time-series data with on-stop fluctuations in each trading session. Forecasting the price trends to minimize the risk of investing can become a guaranteed technique in trading markets [4]. In coffee commodity markets, the amount of coffee transaction data is generated, increasing in volume and variety over the years. This allows data gathering [5], integration, and analysis demands on a scale not seen earlier. Some reports have been published and provided as coffee world trade statistics such as the report Coffee: World Markets and Trade [6], Coffee Market Report [7], Global Market Report: Coffee [8], World Bank Commodities Price Data (Pink Sheet) [9], etc. Typically, these reports will provide readers with up-to-date information on coffee world markets and various units with pre-calculated formulas.

This large amount of data, when properly managed and processed, has the potential to provide accuracy, variety, and speed of analysis to support decision-making and build predictive models. The tasks of collecting data from different types of reports and processing them before consumption can be more convenient, thanks to the application of engineering technology. Advanced Information Communications Technology (ICT) and data mining technologies have the potential to transform the operations of the trading market. Coffee traders can benefit from a thorough analysis of daily and historical market data, enabling them to make more informed trading decisions. However, current systems face specific limitations, including insufficient data, inadequate documentation for storage, and the need for scalable infrastructure to facilitate big data analytics, such as a data warehouse or data lakehouse. Data warehousing techniques evolved as computer systems became more complex and needed to handle this increase. A data warehouse is also a processed database for businesses whose primary purpose is to provide reporting and data analysis [10]. The data warehouse transforms and categorizes data using the Extract – Transform – Load (ETL) process from different data sources before importing it into the repository. This data will be processed to prepare for various consumption purposes such as reporting and analysis.

In the context of coffee trading, collecting transaction data, visualization and application of predictive analysis plays an important role and is also a big challenge for investors. Thereby, a functional system that can collect coffee trading transaction data and apply mining or analyzing techniques to give output as a visualization dashboard in real-time, is a viable idea to address investor needs. Addressing this problem, the paper presents a design and implementation of a coffee commodity trading big data warehouse capable of analyzing various essential parameters for supporting informed decision-making. First, the designed system can automatically collect coffee trading data for New York Arabica coffee futures prices from selected worldwide reports and financial data portals. Next, the ETL process is adopted to ingest coffee futures trading crawled data into the 3 layers data warehouse. Finally, the analytical system will extract and visualize selected key dimensions that influence coffee futures prices within different observation windows and perspectives.

The contributions of this paper can be summarized as the following:

- Provide a design and implementation of a coffee commodity trading big data warehouse capable of analytical various essential trading.
- Build a prototype visualization dashboard and predictive model for supporting informed decision-making.

The remaining paper is organized as follows: Section II provides an overview of previous research or studies related to mining and analysis of the coffee trading market and technology related to building recommendation model; Section III discusses the methodologies for collecting coffee trading data, research design, processing data, building data model and analyzing, mining coffee data; Section IV presents the results of prototype visualization dashboards and discussions; Section V, the last chapter, presents the conclusions and future work to improve the results.

II. RELATED RESEARCH

In this section, we will lay out the theoretical background for the research. The concepts of the coffee derivatives market and an overview of data Warehouse and OLAP are defined. The current state of related works in those areas is analyzed regarding coffee trading markets.

A. FUTURE COFFEE COMMODITY TRADING DATA

Coffee is a popular beverage and an important commodity in the global economy. Coffee has become the product with a high level of consumption worldwide in recent years, with over 2.25 billion cups of coffee consumed daily [4]. Coffee international trades occur between exporting and importing parties, maybe with or without broker parties [11]. Since the 1960s, the international market has been guided by International Coffee Agreements (ICAs) hosted by The International Coffee Organization (ICO) [12]. ICAs contribute to developing a sustainable world coffee sector and reducing poverty in developing countries. There are coffee exchange markets around the world, such as the London Robusta Coffee Market (ICE Futures Europe), the Arabica Coffee New York Market (ICE Futures US), the Builders Merchants Federation Market (São Paulo, Brazil), etc.

The advent of the derivatives market is one of the significant financial innovations in developing the financial system. In the conditions of the modern economy, financial innovations are growing, expanding, diversifying, and creating conditions for the derivatives market to grow. The derivatives market is the place where derivatives are signed and traded. A derivative product is a financial product that results from another product, also known as a base asset (the base asset

may be a bond, stock, currency, or goods) [13]. The value of derivative assets is determined based on the value fluctuations of the base asset. Derivative financial instruments are used to transfer unwanted risks to partners at risk who are compensated or want to acknowledge such a risk. The derivatives coffee market includes primary contracts, such as forward, futures, option, and swap contracts [14]. In our research, we only focus on the coffee futures contract, which is a legal agreement created on an organized exchange to buy/sell a specific asset at a point time in advance at a specified price. The most traded place for Arabica coffee is the ICE exchanges in New York, while for Robusta coffee, the LIFFE exchanges in London are where the derivative contracts are mostly traded. Future prices listed on exchanges are different, depending on the supply and demand of each market. Commodity Futures Trading Commission (CFTC) is an independent federal agency established by Congress in 1974 that regulates the market for goods in terms [15]. The CFTC reserves the right to establish margin requirements, regulate option contracts, and oversee the business of registered commission trust contracts. CFTC publishes its Commitment of Traders (COT) report to help investors understand market moves. Specifically, the COT reports analyze open interest contracts for future contracts and options in the future markets from Tuesday and published on Friday. The CFTC received data from the companies reporting on Wednesday morning, then edited and verified the data for release on Friday afternoon. The CFTC publishes a Trader's Weekly Commitment report, which reveals the size of positions of commercial and non-commercial traders. In the context of this research, CFTC reports could be used to understand market price movements.

In the past, several studies used price data and applied several different algorithms to predict coffee prices. Quarterly data from 1962(1) to 2001(1) coffee price data in the New York market taken from the International Coffee Organization (ICO) were forecasted by using estimated linear and non-linear error correction models [15]. In this paper, they studied the spot prices of four different coffee types: Unwashed Arabica, Colombian Mild Arabica, Mild Arabica, and Robusta. Monthly Indigenous People Robusta price coffee data from January 1995 to September 2015 was used to build time series statistical modeling to forecast prices using the GARCH model, Hybrid ARIMA-ANN Model [16], [17]. Empirical analysis of daily price data of nearby ICE contract and LIFFE contract contracts obtained from Quantl (2016) from 1/1/2006 to 30/6/2016 is used to evaluate co-movement between the futures price and forecast for the Value-at-Risk [18]. Coffee Commodity Price data from Ethiopian Commodity Exchange (ECX) was used in J48 decision tree classification algorithms to determine the determinate critical factors for the commodity [19]. Monthly average price collected from the Center for Advanced Studies in Applied Economics at the Luiz de Queiroz College of Agronomy (ESALQ) - University of São Paulo (USP), Brazils was used to estimate coffee prices based on an extreme learning machines model [20].

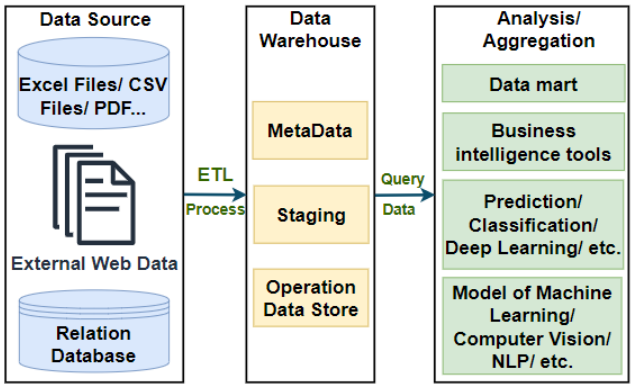

FIGURE 1. Data warehouse architecture.

In previous studies, coffee prices have been used to predict future price trends by multiple models and tested with many different data sets and not over different periods. To the best of our knowledge, although coffee futures trading data, particularly price data, is commonly used, there is no system to store this data. It will bring much research value if these data are stored centrally and uniformly processed, combined with price forecasting algorithms and models. This paper presents a different approach to collecting, storing, processing, and consuming coffee futures transaction data.

B. DATA WAREHOUSE CONCEPT

The term Data Warehouse (DW) was first used as the comprehensive repository of all pertinent data utilized for generating business reports [21]. DW supports complex queries and gathers data from different sources to get complete analysis information. A DW is a set of subject-oriented, complete, leak-free, and historically valuable data. Historically, the database could be understood as a collection of related data representing some real-world element. Along with modern technology, the DW was developed as an information system that stores historical and commutative data from one or more sources [22].

Meanwhile, Fig. 1 presents the data warehouse architecture. First, the source layer of a DW system can come from heterogeneous data sources, such as initially stored in the relational or legacy database or information system outside. In the second layer - DW Layer, Information is stored in a single, reasonably centralized repository DW. The data warehouse can be accessed directly, but it can also be used to create data marts, partially replicating DW content and designed for specific business departments. A metadata repository stores information about sources, access procedures, data organization, users, data center schema, and more. Data mart is efficient and flexibly accessed in the third analytics layer to generate reports, dynamically analyze information, and simulate hypothetical business scenarios. It also features aggregated information navigation, a complex query

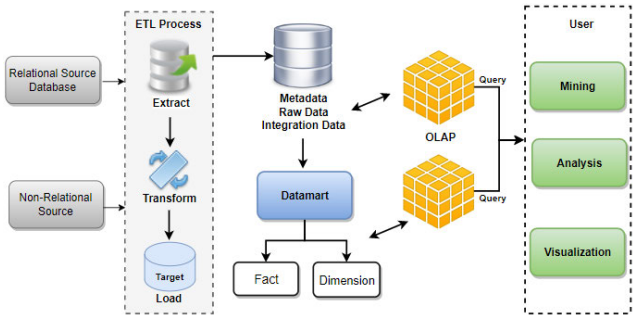

FIGURE 2. OLAP server operations.

optimizer, and customer-friendliness through business intelligence tools, prediction, classification, machine learning, and deep learning techniques.

The DW will help significantly increase the data volume that needs to be aggregated, stored, and processed. DW architecture is based on building a unified data warehouse from many different data sources to serve a query [23] consisting of three main classes: Data Source Class, Warehouse Class, and User Class. The structure of DW can be divided into three main levels. The first level is the physical level, in which all the data is stored, along with metadata and logic processing for data filtering, organizing and packaging, and processing detailed data management [24]. The second is Logical, which contains metadata that includes business rules and processing logic for filtering, organizing, encapsulating, and processing data [24]. Furthermore, the final is an intelligent level or themed data (Datamart) [24]. DW helps to save time and increase query performance and data access when integrated with other data processing such as Online analytical processing (OLAP), which is a part of the decision-making support system [25], is used to optimize for querying and reporting by multiple perspectives dimensions on many aspects of a problem with a large amount of data. OLAP is a computing method that efficiently executes multi-dimensional analytical queries by utilizing historical data and aggregating it into structures for sophisticated analysis [26]. OLAP cube architecture describes data that is represented in a multi-dimension cube. Each dimension describes a particular feature of the data [27].

With such an architecture, it is possible to eliminate the linking between worksheets, reducing performance in the information extraction process. The illustration of OLAP server operations architecture is shown in Fig. 2. The data source, which can be a relational source database or a non-relational source, will be collected and processed in the staging area by the ETL process (which will be more carefully reviewed in the next section). Processed data will move to and be saved in a data warehouse, such as Metadata, Raw Data, and Integration data in fact and dimension. These data will be arranged in the data mart and used in the OLAP server for user mining, analysis, or virtualization.

Performing analytic queries on many relational databases takes a long time because the computer system searches through many data tables. In the OLAP system, precalculated aggregations and integrated data can help generate reports faster and precisely when needed.

OLAP and data mining are used in research [28] to find long-term and short-term rules that affect the underlying time series of oil prices in the data warehouse and make dynamic risk decisions based on usage patterns, relationships, and information. Therefore, a data warehouse was built to store pre-processed and aggregated multidimensional data as a form of OLAP analysis. XML document warehouses for storing unstructured information as a relevant-cube model was presented in [29]. Their model can retrieve documents and facts relevant to the selected context. Each fact will be linked to a set of documents describing its context and assigned a value indicating its relevance to the specified context.

The paper [30] proposed a new design scheme for the financial virtual experiment teaching system and introduced the implementation of the financial data warehouse based on MATLAB and data warehouse. Taiwan and Hong Kong stock markets have strong associations with both inside and outside factors, which was proven in research [31]. Authors used star schema to store different market indices, such as performance, observation date, frequency, etc. They also conducted their analysis with different techniques, such as association rule-based cluster analysis with the k-means algorithm. In research [32], [33], authors indicated individual dimensions with the historical data of that dimension and market indices to analyze the impact of every cuboid by aggregating or integrating these data.

C. DATA LAKEHOUSE DEVELOPMENT

In 2011, the term Data Lake was introduced by Pentaho CTO James Dixon to refer to a large pool of data in its natural, unstructured form [35]. Fig. 3 below shows the basic architecture of the data lake. In data lake architecture, raw data is stored in its original format. After measuring, users can transform, classify, or analyze different pieces of data based on their needs, which need to be further processed when required. In Data Lake, all data (raw data) from system sources are stored, including data sources that may be denied storage in the Data warehouse, such as web server logs, sensor data, social media activity, text, images, etc.

Data lake relies on low-cost storage options, a differentiator from data warehouse technology. While in a data warehouse, data must be cleaned and extracted uniformly, regardless of the data source. Data warehouse and Data Lake differences include data structure, schema appearance, flexibility, adaptability, and performance. Raw data, unstructured or semi-structured, is stored in data lakes, which require a larger storage capacity and are ideal for machine learning. Data lakes have less organization and filtration than data warehouses, as raw data flows with no fixed purpose. Data lakes can be difficult to navigate for those unfamiliar with raw data and require specialized tools for processing and translation. The current development of a data warehouse has

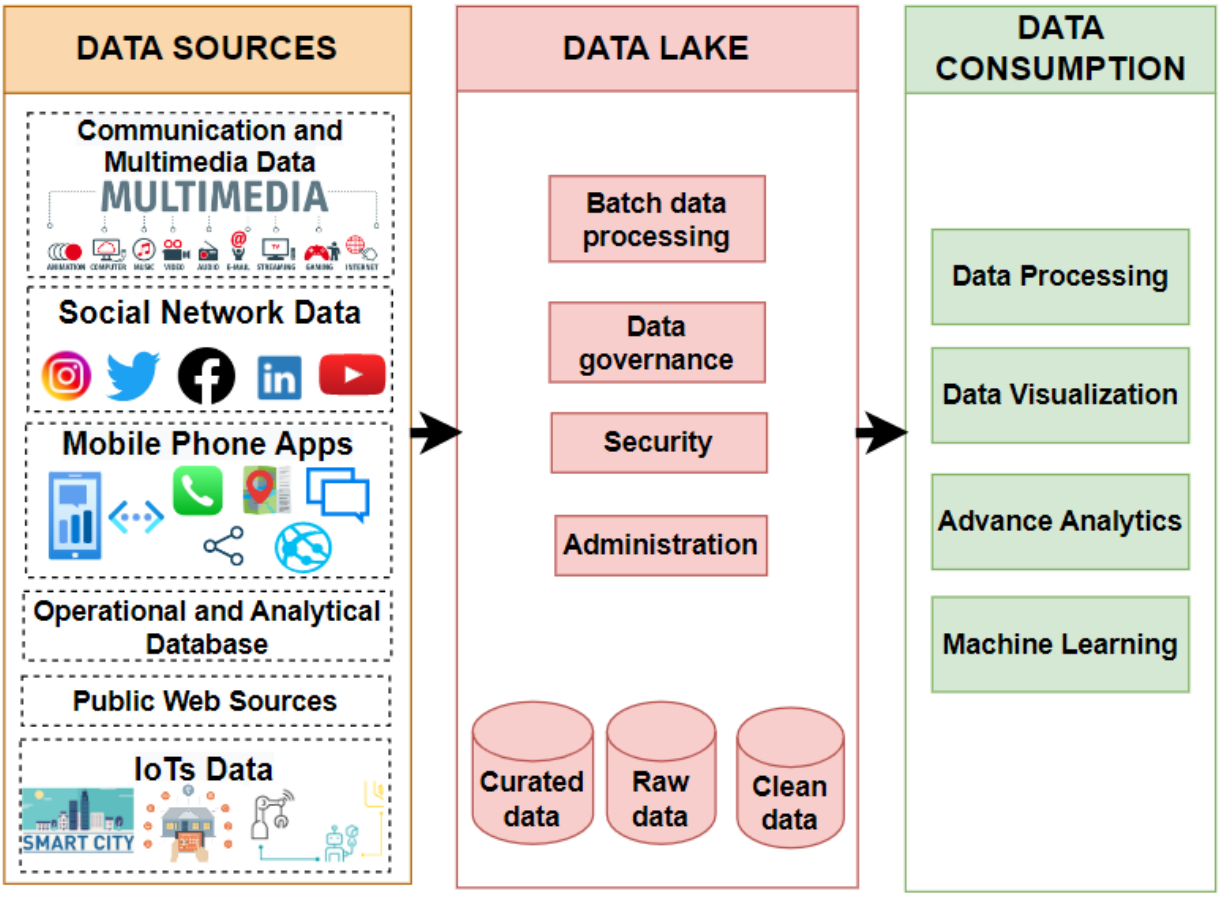

FIGURE 3. Data lake architecture.

evolved into a data lakehouse, which can accommodate the diversity of big data, including structured, text, unstructured, IoT, etc.

However, the research focused on detailed daily price data and parameters related to price data on coffee futures trading. For this data type, the volume will increase rapidly daily but remain the same in diversity. Therefore, data source diversity remains the same throughout the data collection process. This leads to the storage method in traditional databases, such as data warehouses capable of storing structured and transformed data, which has met the research needs. In a data warehouse, stored data must undergo extraction, transformation, and processing through the ETL process before being imported into the archive. In addition, a data warehouse uses processed data, which is used for specific purposes within the organization.

D. PROCESSING DATA IN A DATA WAREHOUSE

In DW, processing data tasks are usually conducted with Extract – Transform – Load (ETL) or Extract – Load – Transform (ELT). The raw data must be cleaned, enriched, and transformed before being integrated into an analyzable whole. Both processes aggregate, organize, and move data to the target system for storage and analysis. Extracting is the process of reading data from a database. During this stage, data is often collected from a variety of sources.

Transforming is converting the data extracted from the previous process into a form that is required to be placed in another database. Conversions happen by using rules, lookup tables, or by combining this data with other data. In the Extracting stage, data from different source systems can be extracted into the staging step in different formats, such as relational database (RDBMS), NoSQL, XML, and flat file [35]. First and foremost, extracting data from different source systems and not loading the data directly into the repository but storing it in staging is paramount [36]. This is because the extracted data has many different formats and may be omitted (data carrying a null value). Hence, downloading it directly into the data storage can cause errors that affect it, and recovery will be much more difficult. Data transformation has rules or functions applied to the extracted data to convert it into a single standard format [32]. It can be related to the following processes or tasks: Attribute filter - loading only specific properties into the data vault; Clear data - fill the NULL values with some default values; Join - combine multiple attributes into one; Split - split a single

attribute into multiple attributes; Sort - sorts tuples based on several attributes (usually by primary key). Loading data is the task that sends data to the DW system. Data is updated by regularly uploading to the data warehouse or executed after set intervals. Speed and load times depend only on the requirements and vary between systems. Because of the relational structure of the OLAP data warehouse, ETL is often used because the data transformation is performed before data transmission. Specifically, first, ETL extracts data from homogeneous or heterogeneous data sources. Next, it sends the data into a staging area. From there, the data goes through cleaning, enrichment, transformation, and finally, stored in the data warehouse.

There are three main approaches to modeling ETL processes: modeling based on mapping expressions and guidelines, modeling based on conceptual constructs, and modeling based on the UML environment [37]. The relational structure of the OLAP data warehouse requires a database schema to reflect the logical relationships between objects and information in data tables such as Star schema or Snowflake schema. The star schema consists of a fact table that can be joined to several normalized dimension tables. When connected, the dimension table will explain information for the fact table. The snowflake schema is equivalent to a star schema. In this schema, the data table is normalized to the dimension tables, which are connected to the child tables. Users benefit from its low data redundancy, but it comes at the expense of query performance.

III. DESIGNING AND IMPLEMENTATION OF COFFEE COMMODITY TRADING DATA WAREHOUSE

A. DESIGNING EXPERIMENT FRAMEWORK

Fig. 4 presents the proposed research model with three main phases: Data source, Data warehouse, and Data consumption. First, data is collected from external web sources. External web data with a relational database and internal files will be arranged and saved in a raw data folder. In the staging class, raw new or unqualified data will be cleaned with Python code. Raw data will also extract historical data from the separated files. Both cleaning and historical data will move to the operation data store class with the 3NF data model, which uses normalizing principles to reduce data duplication, avoid data anomalies, ensure referential integrity, and simplify data management. At this same time, cleaning data will be sent to obtain aggregation data. During the data warehouse processing, the ETL process is continuously extracted from the active database sources, then transformed to fit our data warehouse structure and loaded into the systems.

Data can be stored and processed in three data table types in the data warehouse layer. First, metadata will help to identify data management information. Metadata can make it possible to easily monitor data by identifying the data with its description of all available information and associated work during implementation and monitoring all data using the program and reports [40]. The staging data in our model

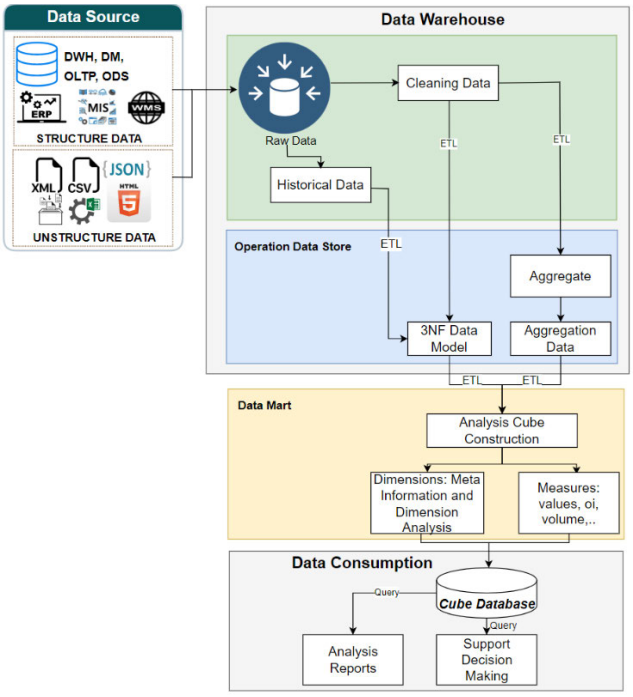

FIGURE 4. Proposed experiment framework.

is data that is extracted 1:1 to ensure data integrity and to store historical data. The encryption of the data dimensions by ID for increased storage efficiency, data processing, and information security is done in the operation data storage (ODS). ODS data will ensure clean and accurate data, according to business rules, and computation and conversion to provide data according to business requirements. ODS data will help to control and record the changes in analysis dimensions from the input data (add, edit, delete).

All qualified data will be sent to Analysis Cube Construction in the data mart class to extract dimensions and to create a dimension table and fact (or value) tables in the Data mart layer. The data mart constructs the dimensions of analysis, constructs measurement data, measures according to business requirements, and builds data models for statistical reporting and analysis on tools or platforms. A data mart will be used directly in the analysis/aggregation class to make analysis reports and to support decision-making. The data mart can also be used to build prediction or classification, deep learning models, and applied machine learning models, such as computer vision or natural language processing.

The implementation process of building the data warehouse in our research is shown in Fig. 5. The first phase is also the focus of this report, demonstrated in the server production process, including data integration (ETL), building a data warehouse, and using the data warehouse to build reports analysis. The data warehouse can be used to build machine learning models, such as model computer vision and natural language processing, to mine and predict effectively.

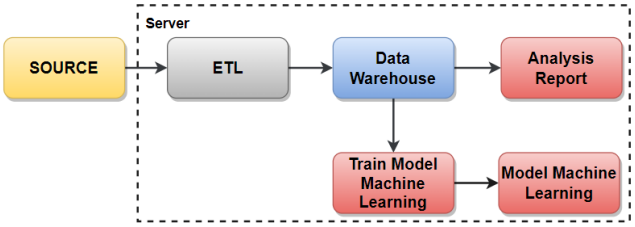

FIGURE 5. Block diagram of implementation process.

B. RESEARCH ENVIRONMENT

In this paper, we build on physical server Intel (R) Xeon (R) CPU E5-2687W v4 @ 3.00GHz, RAM 128 GB, 64-bit Operating System, x64-based processor). We build crawling, cleaning, and processing data programs using Python. We also use PostgreSQL–an open-source relational database management system emphasizing extensibility and SQL compliance to store query data. Power BI tool is used for simple formula calculations data modeling, and data visualization.

C. COLLECTING COFFEE FUTURES TRADING DATA

In this research, we collected the futures coffee contract price from the Yahoo Finance portal. These were the adjusted closing prices of the coffee mark (KC = F) traded on the commodity exchange ICE Futures US. In monthly data, the values are given in US dollar cents per pound (0.453 kg). In this research, we will build our demo for the data from January 2000 to October 2022.

Besides that, the Commitment of Traders (COT) report, which measures traders' trading volume in the US futures trading market, is also collected. Commodity Futures Trading Commission (CFTC) publishes a COT report every Friday.

In addition, the Commodity Index Trader (CIT) report, a supplement to the COT report, shows the relationship between "index traders" positions in that supplement and "swap dealer" positions in the Disaggregated COT, whose there are specific differences. We also collected data from the CIT report in our data warehouse system. We use Python and Splash-scrapy to collect data from the coffee market on websites. Splash is a lightweight headless browser that works as an HTTP API. With Splash, we can render JavaScript pages, follow URLs, and extract data from specific HTML paths on each page. First, we define start URLs, the starting address for the spider. We also define the list of URLs corresponding to the domain in the allowed domains. To accomplish this task, we have researched and identified websites that must be crawled manually.

The pseudocode illustrates the function of downloading URL files from data website sources. As a result, data files are saved with the format of the name of the data source and the date of loading the data to distinguish the data files by different dates.

After downloading the data files from the URLs, we built other functions to parse the data according to the respective

Pseudocodes 1: Collecting Data

Input: URL list, destination path.
Output: URL files downloaded from trading organizations.

```
def download_url_file(dest_path, url_list):
  for url_file in url_list:
    if 'zip' in url_file:
      r = requests.get(url_file, stream =True)
      z = zipfile.ZipFile(io.BytesIO(r.content))
      z.extractall(dest_path)
    else:
      r = requests.get(url_file, stream =True)
        if '!DOCTYPE HTML' not in r.text:
          print(str(datetime.nowstrftime(DATETIME_FORMAT))
+'Download:'+ url_file.split('/')[-1] + 'saved to: ' + dest_path)
            with open(dest_path + url_file. split('/')[-1].lower(),
'wb') as f:
              f.write(r.content)
        if '!DOCTYPE HTML' in r.text:
          print(url_file + 'not found')
```

source. Handling this task, we create a parser, after, read the config file, and then get the section. Here, the default is PostgreSQL. In this source layer, we continue to create a directory to save all the raw data source files to facilitate calling this data and to avoid data loss or error. Since many different data sources exist, we create different sub-folders for each. The data source can be available through internal files (Excel, PDF, CSV files) or related to the current database. After collecting the data, it will be processed with the function inserting to metadata. We create a table staging metadata and put the collected data with a value source code processing stream, a source file location, a type of source file, preprocessing staging, a created date, and a changed date. We checked this current data source and arranged it in the data source files of staging to create consistency in data. External web data with a relational database and internal files will be arranged and saved in a raw data folder. In the staging class, raw new or unqualified data will be cleaned with Python code. Raw data will also extract historical data from the separated files.

D. DATA PROCESSING ETL IN DATA WAREHOUSE

ETL processing contains handling functions that determine precisely where to store the data, provide a function for the driver to collect data from the website, and process the collected data from the website. There are three sub-functions, including (i) Function determines the location to store the data, (ii) Function provides drivers and (iii) Function processes collected data.

ETL process concludes in three stages, as follows. Stage Staging - Five standard functions apply to all process data. Script process staging data, which is function init_db(), and creates a connection to the database; function get_meta_data() gets information metadata for processing; function insert_into_staging() processes and inserts data from raw files for staging; function checking_logs() saves the results of processing; and function main() is the main

process function. Next is the stage of Operation Data Store (ODS). First, we determined all dimensions of the analysis, according to the analysis indexes of exchanges or from data sources pushed into staging. We also built a Slow Changing Dimension (SCD) mechanism to control the change of these analytical dimensions. Next, we created an ODS dimension table. Then, we built ODS fact tables. The construction principle is selecting (filter) and extracting (extract) fields to be calculated, determining the analytical dimensions to bring back the basic ID (integer), basic transformation (add, subtract, multiply, divide, LAG, etc.) on the fields (if any) from staging into ODS. We also inserted ODS Metadata to manage all the information.

Finally, we will continue with stage Datamart. In this stage, based on business needs and the data to support business departments' decision-making, determine the tables to be aggregated (fact) and the fields to be calculated (measures). We created Data Mart Dimension to load data 1:1 from the ODS and select data with four main functions: DB connection, Insert into Datamart Dimension table, process the information to push into the Datamart Dimension table, and the primary function handles. We also inserted a Datamart Metadata to manage all the information. In our system, we also built an Airflow ETL framework to automatically manage all ETL jobs, which helps organize the jobs into directed workflow graphs, monitor them, and keep track of the ETL level activity status. Airflow allows writing custom plugins for databases that are not supported out of the box. Airflow provides a Directed Asymmetric Graph (DAG) view that helps us manage task flow and serves as documentation for many jobs. We also have an easy-to-use web interface that facilitates monitoring and managing ETL jobs. Airflow uses PostgreSQL to store the configuration and status of all DAGs and running tasks.

IV. RESULT AND DISCUSSION

A. FUTURE COFFEE TRADING DATA WAREHOUSE

Information is stored in a reasonably centralized data store. The data warehouse can be accessed directly, but it can also be used to create data marts, partially replicate data warehouse content, and be designed for specific business departments. A metadata repository stores information about sources, access procedures, data organization, users, and data center schema. Dimension tables describe the category by which the numerical data in the block are divided for analysis. When specifying a dimension, we identify one or more columns of these dimension tables. When choosing complex columns, these tables must be related to whose values can be organized in a single hierarchy. To reach the target of defining a hierarchy, we sort columns from most general to most specific. Each column in the dimension contributes to a level for the dimension. The levels are ordered separately and organized in a hierarchy that assumes logical paths. Analytical dimension tables are presented in blue tables in Fig. 6. The data model is designed to track coffee futures price data over time and

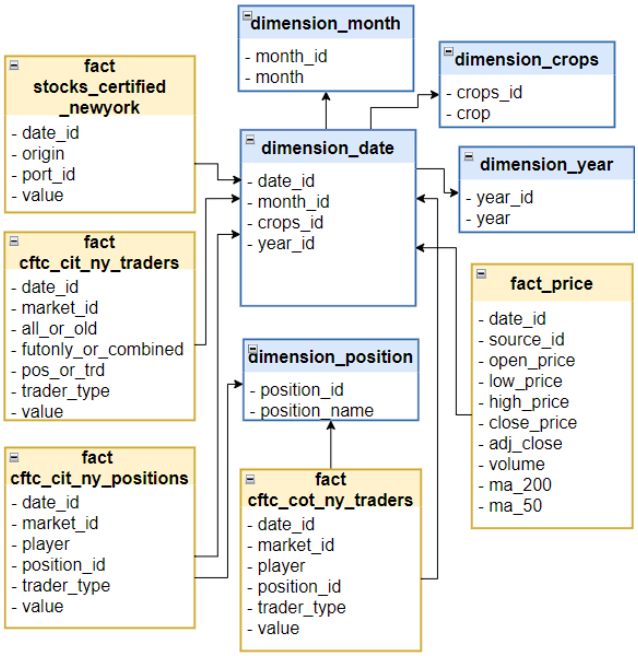

FIGURE 6. Future coffee trading data warehouse schema.

changes in trading positions. To accomplish this goal, the dimension_date table with the sub-dimension tables month, crop, and year. In addition, the position analysis dimension has also been created in the dimension_position table to track trading positions. The COT and CIT reports separate the market participants into traders, brokers, manufacturers, significant speculators, hedge funds, etc. Commercials are the major businesses that use currency futures to hedge and protect them from excessive changes in exchange rates. Non-commercial are individual traders, hedge funds, and financial institutions. These are mostly speculative traders. Reportable Positions present the total number of options and futures contracts that must be reported to the CFTC. Non-Reportable Positions present are the number of unfilled contracts that do not meet the reporting requirements of the CFTC. Analyzing the trade change in the position's direction will bring the trader much analytical value.

In value tables, we define two types of columns, which contain numeric facts (often called measures) and contain keys to dimension tables. The facts table contains granular-level events or events that have been aggregated. Fact tables that contain aggregated facts are often called summary tables. The facts table usually includes events with the same level of aggregation. However, most events are always binding, they can be associated with some dimensions or not. There are nine fact tables used as values in our data warehouse. Fact tables consist of a business process's measurements, metrics, or facts. They are located at the center of a star schema or a snowflake schema, surrounded by dimension tables. All the key and values fields are presented as yellow tables in Fig. 6. Each value table contains dimension key columns related

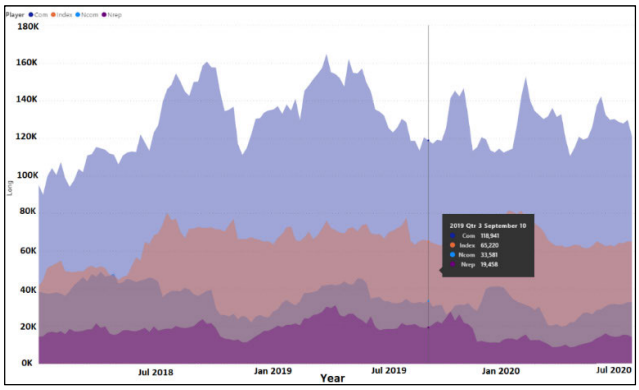

FIGURE 7. Summarize buy coffee futures contract.

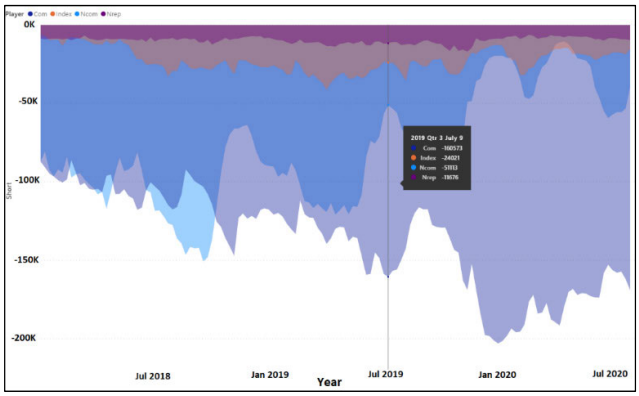

FIGURE 8. Summarize the sale coffee futures contract.

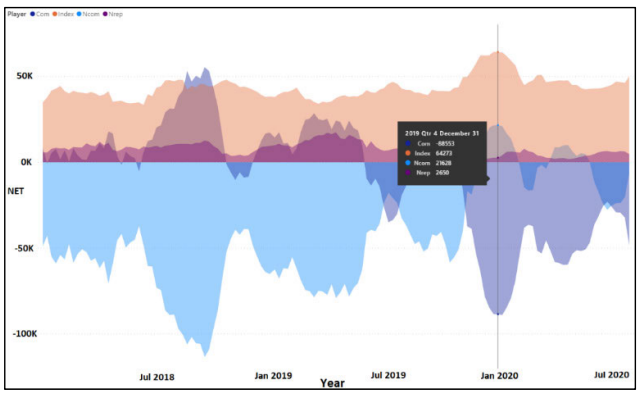

FIGURE 9. Difference between buy/sell in coffee futures contract.

FIGURE 10. MA of arabica prices in the us trading market.

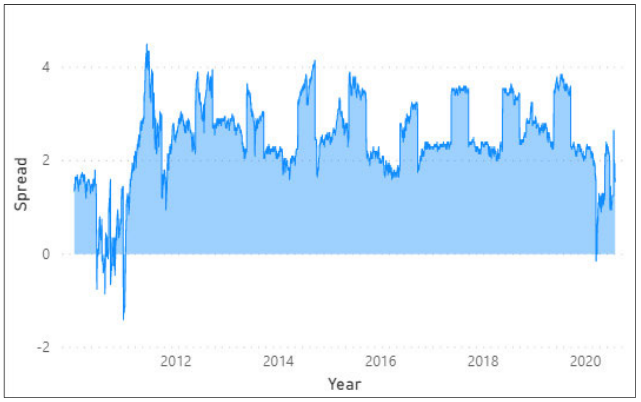

FIGURE 11. Spread fluctuations in the US trading market.

to the dimension tables and columns containing numeric measure data. Numeric measure data built in our system are New York coffee trading price value of trader types in the exchanges.

As part of the research's requirement, it is necessary to design the data structure for the coffee trading market data mart. This research analyzes fluctuating prices based on the trading positions and players of the New York future coffee market. For the design of this data mart, we used the snowflake schema and added as much granularity as possible to foster better analysis. Based on the dimensions and facts identified, we introduce a model data warehouse schema for coffee commodity trading exchanges in a logical view in Fig. 6.

B. BUILDING DATA WAREHOUSE REPORT (VISUALIZATION) THE VOLATILITY OF POSITIONS IN THE NEW YORK EXCHANGES

We will present some visualization dashboards built in Power BI, a business analytics service provided by Microsoft [38], which can help to build interactive visualizations and business intelligence capabilities.

In our work, we created three visualization charts to describe the changing COT report time by time. Long is the total number of contracts to buy reported to the CFTC. Short is the total number of short contracts reported to the CFTC. Open interest is the number of unexecuted contracts. Number of traders is the total number of transactions required to be reported to the CFTC. Based on their behavior changes in buying contracts, investors can predict the upcoming market trending the market's demand. Fig. 7 presents the total number of purchase contracts in the US futures trading market. The volatility of the purchase contracts is presented as four types of investors over the years. Investor Commercial has the largest and increasing position volatility, peaking in 2019.

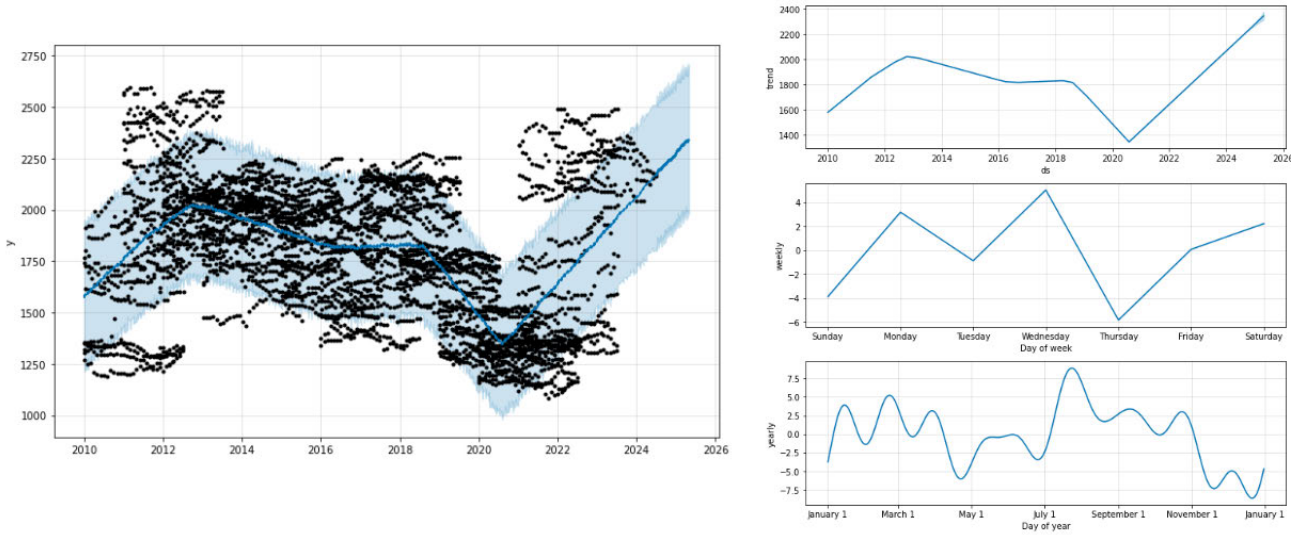

FIGURE 12. New york price seasonal analysis.

As the year 2020 was affected by the COVID-19 pandemic, there was a slight decrease. Other investors' changes can be easily tracked in our chart.

Fig. 8 presents the total number of sale contracts in the US futures trading market. The volatility of the sale contracts is presented as four types of investors over the years. The COT chart also separates market participants, including Investors, brokers, manufacturers, large speculators, and hedge funds. These components all play a crucial role in the Commodity Market. Based on their behavior changes in sale contracts, Investors can predict the upcoming market trending the market supply. Other investors' sales changes can be easily tracked in our chart.

Net position is the most important part of the COT report. The net position represents the net buying and selling amount of market participants and is calculated as follows:

Net position = Long Position – Short Position

Among the three market participants groups, the one closest to the price is the speculative group (traders). However, in some moments, there are also differences, such as speculators falling but prices are still rising or speculators rising but prices are still rising when prices are falling. Moreover, these special times can bring helpful information to evaluate the market.

Fig. 9 presents the Difference between Buy and Sell in Coffee Future Contracts in the US futures trading market. Users can also recognize a shortage of supply or demand in the coffee trading market. When the net position increases, the demand increases, and the probability of price increases. The net Position line is upward, the trend is bullish, and vice versa.

A moving average (MA) represents volatility and indicates the direction of futures prices over a period. The main purpose of the MA is to track whether the price is moving in an uptrend, downtrend, or without a trend. The MA is considered a slow indicator. It does not affect forecasting but will mainly move according to the established price movements. Moving averages will have a certain lag from price (especially in the short term). We have chosen 50 days for the medium term and 200 days for the long term to model and track as moving averages for the coffee exchange market. Our charts also indicate a separate line for tracking New York prices. The chart in Fig. 10 shows the correlation between the medium-term moving average and the long-term moving average with the actual trading price of coffee in the market. When the short-term MAs make a downward cross with the long-term lines, this can signal a downtrend. When the short-term MAs make a downward cross with the long-term lines, this can signal a downtrend. In November 2019, there was a cross between MA 200 and MA 50. It is a signal of a downtrend price, and the price was down.

Spread plays an important role when tracking the coffee market prices. We also created a chart to track this index. Fig. 11 presents dashboard tracking spread fluctuations in the New York market. Spread is the difference between the bid and ask prices. In other words, the spread represents the difference between the demand and supply prices of coffee.

Many factors affect bid-ask spread on certain exchanges. High volume is a sign of a highly liquid market. This means the bid and ask prices will be close together, resulting in low spreads. Conversely, if the trading volume is small, it indicates low liquidity, the bid and ask prices are far apart and the spread is high. Countries with chaotic political institutions or unstable economies often have relatively high inflation rates and ineffective monetary policy. This causes sellers to treat the currency as a risky investment and sell at a higher price. Meanwhile, buyers always want to buy at a discount to cover up. This causes the spread to be spaced. A currency not backed by a tight monetary policy and a stable central bank

is often prone to significant volatility. As a result, sellers will push the ask price higher, resulting in a widening spread.

We built a demo applying the Prophet model to analyze the seasonal parameters of the data. We can see trends daily, weekly, and yearly by using forecasting time series data based on an additive model where non-linear trends fit with yearly, weekly, and daily seasonality. We hope this works for coffee price data, which has the characteristic of a time series with strong seasonal effects and several seasons of historical data. In Fig. 12, the trend of New York prices is predicted. We build different charts to visualize the analysis. We build charts to analyze the trend of price by the year, day of the week, and day of the year.

Our data warehouse is designed for data scalability. The data structure aggregates the measurements according to the levels and hierarchies of each dimension to be analyzed. The blocks combine several dimensions and can hyper-variability of the concept of complexity. In our system, integrated data is efficiently and flexibly accessed to generate reports, dynamically analyze information, and simulate hypothetical business scenarios. One drawback of this architecture is that additional file storage space is used through the redundant conditioning layer. It also makes the analytics tools further away from real-time.

V. CONCLUSION

Coffee, the second-largest global soft commodity, can take advantage of a comprehensive mining of daily and historical market data for more effective informed trading decisions. The paper presents a design and implementation of a coffee commodity trading Big Data warehouse supporting informed decision-making for trading coffee of small and mid-size enterprises (SMEs), business households, farmers, and others concerned through the visualization analysis dashboards. Specifically, we collected data from different sources, including the coffee market data from the New York Market of Intercontinental Exchange (ICE) from the portal Yahoo Finance, the Commitment of Traders (COT) report, and the Commodity Index Trader (CIT) report from January 2000 to October 2022. The Extract, transform, and load (ETL) process is adopted to ingest coffee futures trading crawled data into the 3 layers data warehouse. Finally, the analytical system will extract and visualize selected key dimensions that influence coffee futures prices within different observation windows and perspectives. Our research focused our visualization on four main parameters: market price, volume, trailing over 250 days, and the highest/lowest 40 days. As a result, we implement a prototype of a coffee trading data warehouse on the crawled data and visualize trends in coffee futures prices.

Future work will be needed to collect data from the other markets and compare the correlation between different markets. We also want to define more parameters that can affect the coffee price. We tend to build an Android app based on cloud computing and use model machine learning, model computer vision, or natural language processing to build a real-time prediction application. The market for derivatives develops in parallel with the need to manage price changes of critical commodities. According to modern financial theory, introducing a derivative market in a spot market helps to complete the process of market-based pricing and prevent the risk of price fluctuations in the spot market at a low cost. In research [39], [40], authors synthesized a complete case study of commodity futures and agricultural option markets regarding the inter-time price relationship of stocks, speculation, price behavior, and institutional problems. The authors raise controversial issues and challenges for the studies focusing on risk management and market strategy, price and volatility behavior, e-transactions and trading funds, and international framework. In the coffee derivatives market, these problems will be affected by similar effects and problems. In the future, we will focus on data analysis and discovery and build effective prediction models for the coffee derivative market based on deep learning or machine learning algorithms.

AVAILABILITY OF DATA AND MATERIALS

Please contact the corresponding author for data requests. The coding environment employed is Visual Studio, encompassing Python programming in ETL Processes. The database management system in use is PostgreSQL. The authors have published the resources of the article on GitHub [40].

REFERENCES

[1] S. Ponte, "The 'latte revolution'? Regulation, markets and consumption in the global coffee chain," *World Develop.*, vol. 30, no. 7, pp. 1099–1122, Jul. 2002.
[2] *Coffee Industry Research Report*, Agricult. Rural Develop. Dept., Int. Bank Reconstruct. Develop., New York, NY, USA, 2018.
[3] T. T. K. Hong, "Effects of exchange rate and world prices on export price of Vietnamese coffee," *Int. J. Econ. Financial*, no. 6, pp. 1756–1759, 2016.
[4] M. C. Angadi and A. P. Kulkarni, "Time series data analysis for stock market prediction using data mining techniques with R," *Int. J. Adv. Res. Comput. Sci.*, vol. 6, no. 6, 2015.
[5] M. Obitko and V. Jirkovský, "Big data semantics in industry 4.0," in *Proc. 7th Int. Conf. Ind. Appl. Holonic Multi-Agent Syst. (HoloMAS)*, Valencia, Spain. Springer, Sep. 2015, pp. 217–229.
[6] *Coffee: World Markets and Trade (Report)*, Foreign Agricult. Service/USDA, Global Market Anal. United States Dept. Agricult., Dec. 2022.
[7] *Monthly Coffee Markets Report*, Int. Coffee Org., Apr. 2022.
[8] V. Voora, S. Bermúdez, and C. Larrea, "Global market report: Cocoa. Global market report: Coffee," Int. Inst. Sustain. Develop., Winnipeg, MB, Canada, Tech. Rep., 2019.
[9] World Bank. (2023). *World Bank Commodities Price Data (The Pink Sheet)*. Accessed: 2024. [Online]. Available: https://www.worldbank.org/commodities
[10] P. Paulraj, *Data Warehousing Fundamentals for IT Professionals*, 2nd ed. Hoboken, NJ, USA: Wiley, 2011.
[11] M. T. A. Pieterse and H. J. Silvis, *The World Coffee Market and the International Coffee Agreement*. Agricultural Univ., 1988.
[12] T. M. Leonard, *Encyclopedia of the Developing World*. Evanston, IL, USA: Routledge, 2013.
[13] S. B. Bush, "Derivatives and development: Contemporary applications," in *Derivatives and Development: A Political Economy of Global Finance, Farming, and Poverty*. New York, NY, USA: Macmillan, 2012, pp. 13–49.
[14] P. H. May, G. C. C. Mascarenhas, and J. Potts, "Sustainable coffee trade: The role of coffee contracts," Tech. Rep., 2004.
[15] Commodity Futures Trading Commission. *About Commodity Futures Trading Commission*. Accessed: 2024. [Online]. Available: https://cftc.gov/About/AboutTheCommission

[16] K. Naveena, "Statistical modelling for forecasting of price and export of Indian coffee," Tech. Rep., 2015.
[17] K. Naveena and S. Subedar, "Hybrid time series modelling for forecasting the price of washed coffee (Arabica Plantation coffee) in India," *Int. J. Agricult. Sci.*, 2017.
[18] P. Fousekis, "Price co-movement and the hedger's value-at-risk in the futures markets for coffee," *Agricult. Econ. Rev.*, vol. 18, no. 1, pp. 35–47, 2017.
[19] W. A. Degife and D. I. A. Sinamo, "Efficient predictive model for determining critical factors affecting commodity price: The case of coffee in Ethiopian commodity exchange (ECX)," *Int. J. Inf. Eng. Electron. Bus.*, vol. 11, no. 6, pp. 32–36, Nov. 2019.
[20] C. Deina, M. H. do Amaral Prates, C. H. R. Alves, M. S. R. Martins, F. Trojan, S. L. Stevan, and H. V. Siqueira, "A methodology for coffee price forecasting based on extreme learning machines," *Inf. Process. Agricult.*, vol. 9, no. 4, pp. 556–565, Dec. 2022.
[21] B. A. Devlin and P. T. Murphy, "An architecture for a business and information system," *IBM Syst. J.*, vol. 27, no. 1, pp. 60–80, 1988.
[22] D. Cardon, "Database vs data warehouse: A comparative review," Health Catalyst, South Jordan, UT, USA, Tech. Rep., 2014.
[23] G. Srivastava, J. C. Lin, X. Zhang, and Y. Li, "Large-scale high-utility sequential pattern analytics in Internet of Things," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12669–12678, Aug. 2021.
[24] J. C.-W. Lin, Y. Djenouri, and G. Srivastava, "Efficient closed high-utility pattern fusion model in large-scale databases," *Inf. Fusion*, vol. 76, pp. 122–132, Dec. 2021.
[25] J. Dixon. (Oct. 2010). *Pentaho, Hadoop, and Data Lakes*. Accessed: Jun. 5, 2022. [Online]. Available: https://jamesdixon
[26] C. Adamson, *Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance*. Hoboken, NJ, USA: Wiley, 2012.
[27] A. Kishor and C. Chakarbarty, "Task offloading in fog computing for using smart ant colony optimization," *Wireless Pers. Commun.*, vol. 127, no. 2, pp. 1683–1704, Nov. 2022.
[28] J. Han, "OLAP mining: An integration of OLAP with data mining," in *Proc. IFIP 7th Conf. Database Semantics*, Leysin, Switzerland. Boston, MA, USA: Springer, 1997, pp. 3–20.
[29] J. M. Perez, R. Berlanga, M. J. Aramburu, and T. B. Pedersen, "R-cubes: OLAP cubes contextualized with documents," in *Proc. IEEE 23rd Int. Conf. Data Eng.*, 2007, pp. 1477–1478.
[30] Z. Y. Jiang, "Design and implementation of financial virtual experiment teaching system based on MATLAB and data warehouse," *Appl. Mech. Mater.*, vols. 635–637, pp. 1535–1539, Sep. 2014.
[31] S.-H. Liao and S.-Y. Chou, "Data mining investigation of co-movements on the Taiwan and China stock markets for future investment portfolio," *Expert Syst. Appl.*, vol. 40, no. 5, pp. 1542–1554, Apr. 2013.
[32] D. Mondal, G. Maji, T. Goto, N. C. Debnath, and S. Sen, "A data warehouse based modelling technique for stock market analysis," *Int. J. Eng. Technol.*, vol. 7, no. 3.13, p. 165, Jul. 2018.
[33] S. Hightower, W. Biggs, T. Frazer, P. Krause, and E. Biggs, "System and method of real estate data analysis and display to support business management," U.S. Patent 11 065 893, Sep. 1, 2005.
[34] P. Dhanda and N. Sharma, "Extract transform load data with ETL tools," *Int. J. Adv. Res. Comput. Sci.*, vol. 7, no. 3, 2016.
[35] F. Burstein, C. W. Holsapple, N. Jukic, B. Jukic, and M. Malliaris, "Online analytical processing (OLAP) for decision support," in *Handbook on Decision Support Systems 1: Basic Themes*, 2008, pp. 259–276.
[36] M. M. Singh, "Extraction transformation and loading (ETL) of data using ETL tools," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 6, pp. 4415–4420, Jun. 2022.
[37] S. Vyas and P. Vaishnav, "A comparative study of various ETL process and their testing techniques in data warehouse," *J. Statist. Manage. Syst.*, vol. 20, no. 4, pp. 753–763, Jul. 2017.
[38] T. Lachev and E. Price, *Applied Microsoft Power BI: Bring Your Data to Life!* Prologika Press, 2018.
[39] N. B. V. Le, "Building coffee commodity trading data warehouse: Decision making support applying big data analysis," M.S. thesis, Dept. Data Inform., Nat. Korea Maritime Ocean Univ., Busan, South Korea, 2022, pp. 1–40.
[40] N. B. V. Le, Y.-S. Seo, and J.-H. Huh. *Building Coffee Commodity Trading Data*. [Online]. Available: https://github.com/vanlnb/Datasource-of-Coffee-data-warehouse

NGOC-BAO-VAN LE received the B.E. degree in management information system from Vietnam National University Ho Chi Minh City, Vietnam, in 2019, and the M.S. degree in data informatics and in coffee big data and the prediction of coffee spot and futures prices from the National Korea Maritime and Ocean University, Busan, Republic of Korea, in February 2022, where she is currently pursuing the Ph.D. degree with the Department of Data Informatics, focusing on research related to commodities, such as coffee and corn.

Since September 2020, she has been a Research Assistant with the Big Data Center for Total Lifecycle of Shipbuilding and Shipping, National Korea Maritime and Ocean University. In September 2022, she has joined Global R&E Program for Interdisciplinary Technologies of Ocean Renewable Energy (BK 21 Four Research Group) of Interdisciplinary Major of Ocean Renewable Energy Engineering, National Korea Maritime and Ocean University. Her research primarily involves predicting spot and futures prices.

Ms. Le received the Best Paper Award at the 15th International Conference on Multimedia Information Technology and Applications (MITA 2019).

YEONG-SEOK SEO (Member, IEEE) received the B.E. degree in computer science from Soongsil University, Seoul, Republic of Korea, in 2006, and the M.E. and Ph.D. degrees in computer science from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 2008 and 2012, respectively.

From September 2012 to December 2013, he was a Postdoctoral Researcher with the Institute for Information and Electronics, KAIST. From January 2014 to August 2016, he was a Senior Researcher with Korea Testing Laboratory (KTL), Seoul. From September 2016 to August 2022, he was an Assistant Professor (tenure track) with the Department of Computer Engineering, Yeungnam University, Gyeongsan, Gyeongbuk, Republic of Korea, where he has been an Associate Professor (tenure track), since September 2022. His research interests include software engineering, artificial intelligence, the Internet of Things, and big data analysis.

Prof. Seo is a member of board of directors of the Software Engineering Society, South Korea. He was a recipient of the Undang Academic Paper Award (grand prize), in 2022; the Second JIPS Survey Paper Awards, in 2019; and the Best Paper Award at ASK 2022, ASK 2021, MITA 2021, and 2019. Furthermore, he is involved in international standardization activities and is a member of the Korean National Body Mirror Committee to ISO on IT Service Management and IT Governance (ISO/IEC JTC1/SC40). He has served as the chair for international conferences and workshops, the Proceedings Co-Chair for APSEC 2018, the Publicity Chair for WITC 2019–2024, and the Program Chair for HCIS 2019. He also served as a Technical Committee Member for some international conferences and workshops, such as ICSE 2020, ICSE 2020 Demonstrations Track, ICSE 2020 Software Engineering in Practice, MITA 2019–2021, QRS 2020, CSA 2020–2023, WITC 2021–2024, FutureTech 2021–2023, CUTE 2021–2023, CSA 2021–2023, ACIIDS 2023, and ICCCI 2023. He is an Associate Editor of *Human-Centric Computing and Information Sciences* (HCIS) (SCI-indexed), *Processes* (SCI-indexed), and *Electronics* (SCI-indexed). He was also a Guest Editor of *Journal of Systems and Software* (JSS) (SCI-indexed). For more details please visit https://scholar.google.com/citations?hl=ko&user=RnV9ZSgAAAAJ.

JUN-HO HUH (Member, IEEE) received the B.S. degree in science from the Department of Major of Applied Marine Sciences, the B.E. degree in engineering (double major) from the Department of Computer Engineering, Jeju National University at Ara, Jeju, Republic of Korea, in August 2007, the M.A. degree in education from the Department of Computer Science Education, Pukyoug National University, Daeyeon, Busan, Republic of Korea, in August 2012, and the Ph.D. degree in engineering from the Department of Major of Computer Engineering, Graduate School, Pukyoug National University, in February 2016. He finished the Cooperative Marine Science and Engineering Program, Texas A&M University at Galveston, USA, in August 2006.

He was a General/Head Professor with the Catholic University of Pusan on International Game Exhibition G-Star 2017 (G-Star 2017). He was a Research Professor with Dankook University, Jukjeon, Yongin, Republic of Korea, from July 2016 to September 2016. Also, he was an Assistant Professor with the Department of Software, Catholic University of Pusan, Republic of Korea, from December 2016 to August 2019. Also, he was an Assistant Professor with the Department of Data Informatics, National Korea Maritime and Ocean University, Republic of Korea, from September 2019 to September 2021. Since October 2021, he has been an Associate Professor (tenured) with the Department of Data Informatics/Data Science, National Korea Maritime and Ocean University. Since September 2020, he has been the Center Chair (Director) of Big Data Center for Total Lifecycle of Shipbuilding and Shipping, National Korea Maritime and Ocean University. Since September 2022, he has been a Join Associate Professor (tenured) of Global R&E Program for Interdisciplinary Technologies of Ocean Renewable Energy (BK 21 Four Research Group) of Interdisciplinary Major of Ocean Renewable Energy Engineering, National Korea Maritime and Ocean University. He is the author of *Smart Grid Test Bed Using OPNET and Power Line Communication* (IGI Global, USA, 2017); and *Principles, Policies, and Applications of Kotlin Programming* (IGI Global, USA, 2023). He has authored/edited ten books and edited ten special issues in reputed Clarivate Analytics Index journals. He has published more than 100 articles in Clarivate Analytics Index (SCI/SCIE/SSCI indexed) with over 3400 citations and has an H-index of 32.

Dr. Huh received the Best Paper Minister Award (Ministry of Trade, Industry and Energy, Korea Government) at the 16th International Conference on Control, Automation and Systems, ICROS with IEEE Xplore, in October 2016. Also, he received the Springer *Nature* Most Cited Paper Award and the Human-Centric Computing and Information Sciences Most Cited Paper Award, in 2019 (research published in the journal from 2016 to 2018; SCIE IF is 6.60). He was awarded the Commendation for Meritorious Service in the Promotion of Busan's Data Industry, in December 2023 (Commendation by the Mayor of Busan Metropolitan City, Korea Government). Also, he was the Organizing Chair of the 15th International Conference on Multimedia Information Technology and Applications (MITA 2019: University of Economics and Law (UEL), Vietnam National University Ho Chi Minh City, Vietnam); and the 17th International Conference on Multimedia Information Technology and Applications (MITA 2021: Jeju KAL Hotel, ROK). Since 2017, he has been a Technical Committee (TC) Member of International Federation of Automatic Control (IFAC), CC 1 (Systems and Signals), and TC 1.5. (Networked Systems); IFAC, Computers, Cognition and Communication (CC 3), and TC 3.2 (Computational Intelligence in Control); IFAC, CC 7 (Transportation and Vehicle Systems), and TC 7.2. (Marine Systems); IFAC, CC 2 (Design Methods), and TC 2.6. (Distributed Parameter Systems), since 2020. Currently, he was the Managing Editor (ME) of *Journal of Information Processing Systems* (JIPS), Korea Information Processing Society (SCOPUS/ESCI-indexed), from January 2020 to December 2021; and *Journal of Multimedia Information System* (JMIS), Korea Multimedia Society (EI/KCI-indexed), from January 2017 to December 2022. Since July 2023, he has been the Editor-in-Chief (EiC) of *Journal of Global Convergence Research*, Global Convergence Research Academy, Pukyoug National University, Daeyeon, Republic of Korea. Also, he is an Associate Editor (AE) of *Journal of Information Processing Systems* (JIPS), Korea Information Processing Society (SCOPUS/ESCI-indexed); *Journal of Multimedia Information System* (JMIS), Korea Multimedia Society (EI/KCI-indexed); and *Human-Centric Computing and Information Sciences* (HCIS) (SCIE IF is 6.60). For more details please visit https://scholar.google.co.kr/citations?user=cr5wjNYAAAAJ&hl.