Modelling of a business intelligence system for indicator management in the *Stirling* tilapia farming

Modelado de un sistema de inteligencia de negocios para la gestión de indicadores en el cultivo de tilapia *Stirling*

GONZÁLEZ-RODRÍGUEZ, Sabino†, CHÁVEZ-VALDEZ, Ramona*, ARCEO-DÍAZ, Santiago and BRICIO-BARRIOS, Elsa

Tecnológico Nacional de México, I.T. Colima, Division of Graduate Studies and Research, Colima, Mexico.

ID 1st Author: *Sabino, González-Rodríguez* / **ORC ID**: 0000-0002-5008-5984, **CVU CONACYT ID**: 1154657

ID 1st Co-author: *Ramona, Chavez-Valdez* / **ORC ID**: 0000-0002-5697-6825, **CVU CONACYT ID**: 435574

ID 2nd Co-author: *Santiago, Arceo-Díaz* / **ORC ID**: 0000-0002-7085-3653, **CVU CONACYT ID**: 231269

ID 3rd Co-author: *Elsa, Bricio-Barrios* / **ORC ID**: 0000-0002-1260-9740, **CVU CONACYT ID**: 509944

DOI: 10.35429/JCSI.2022.22.8.23.35 Received July 30, 2020; Accepted December 30, 2020

Abstract

The article presents the modeling of a system for the management of indicators during the farming of Stirling tilapia (Oreochromis niloticus Stirling) based on business intelligence, supporting decision making to improve production efficiency. Typically, tilapia production is based on three stages: breeding, rearing, and farming, where a significant volume of data is recorded. A Stirling Tilapia farm in Manzanillo, Colima, Mexico is analyzed. This farm has C-TRA, a Web system that stores data based on the ISO 12877:2011 standard. However, there is a lack of tools to analyze the records. As a proposed solution, an extraction, transformation and loading (ETL) process is used on the original database, a data warehouse model is generated with a star-shaped schema, considering a table of facts and six dimensions, where physicochemical parameters, the amount of feed supplied, and medications are related. In addition, the use of statistical methods Data Envelopment Analysis (DEA) and bootstrap is expressed, in order to obtain the productive efficiency indicators. This design exposes an incremental product innovation in the technological area of the aquaculture sector, and its integration to a data management system, facilitating the study and decision making.

Aquaculture, Business intelligence, Data warehouse, Information systems

Resumen

El articulo presenta el modelado de un sistema para la gestión de indicadores durante el cultivo de la tilapia Stirling (Oreochromis niloticus Stirling) basado en la inteligencia de negocios, apoyando la toma de decisiones para mejorar la eficiencia productiva. Normalmente, la producción de tilapia se basa en tres etapas: reproducción, hormonización y cultivo, donde se registra un volumen importante datos. Se analiza una granja productora de Tilapia Stirling en Manzanillo, Colima, México. Esta granja cuenta con C-TRA, un sistema Web que almacena datos basado en la norma ISO 12877:2011. No obstante, se carece de herramientas que analicen los registros. Como propuesta de solución, se emplea un proceso de extracción, transformación y carga (ETL) sobre la base de datos original, se genera un modelo de almacén de datos con un esquema en forma de estrella, considerando una tabla de hechos y seis dimensiones, donde se relacionan parámetros fisicoquímicos, la cantidad de alimento suministrado, y los medicamentos. Además, se expresa el uso de los métodos estadísticos: Análisis Envolvente de Datos (DEA) y bootstrap, con el fin de obtener los indicadores de eficiencia productiva. Este diseño expone una innovación incremental de producto en el área tecnológica del sector acuícola, y su integración a un sistema de gestión de datos, facilitando el estudio y toma de decisiones.

Acuicultura, Almacén de datos, Inteligencia de negocios, Sistemas de información

Citation: GONZÁLEZ-RODRÍGUEZ, Sabino, CHÁVEZ-VALDEZ, Ramona, ARCEO-DÍAZ, Santiago and BRICIO-BARRIOS, Elsa. Modelling of a business intelligence system for indicator management in the *Stirling* tilapia farming. Journal of Computational Systems and ICTs. 2022. 8-22:23-35.

* Author Correspondence (E-Mail: pablobrimed@gmail.com)
† Researcher contributing as first Author.

Introdution

Aquaculture is the farming of aquatic organisms in both coastal and inland areas, involving interventions in the rearing process to increase production (FAO, 2019). In Mexico, since 1923, aquaculture has been defined as the use of waters and rivers for the breeding and reproduction of animals for own consumption. In addition, since the 80s, aquaculture production began to focus on commerce, placing itself in the primary sector as the most developed food producer (INAES, 2018).

Due to its importance, global aquaculture productivity has increased since the late 90s, with an approximately production of 10 million tons annually, conversely, in 2018 was produced around 114.5 million tons (FAO, 2020).

Meanwhile, in Mexico, aquaculture production increased considerably from 2010, where approximately 10 thousand tons were generated, maintaining a stable volume of 54 thousand tons per year between 2014 and 2018. In 2018, Chiapas was the largest producer in the country, where about 25 thousand tons were reported, followed by Jalisco, Veracruz, and Tabasco. On the other hand, Colima ranked 17th in Mexico, contributing 329 tons of tilapia, 0.62 percent of national production (CONAPESCA, 2021).

Among the factors for the increase in aquaculture production, can be mentioned the use of diverse and innovative technologies that have an impact on the optimization of processes, the reduction of water consumption and the search for a lower use of certain inputs (Ipac, 2021).

Likewise, the use of statistical and administrative techniques makes it possible to generate a projection of current and historical data, providing performance references and productive trends, with the purpose of supporting preventive or corrective actions. Therefore, a useful tool is Business Intelligence (BI), a process of collecting, storing, and analyzing business operations data, which provides comprehensive metrics, almost in real time. Furthermore, it helps to act appropriately, based on a combination of business analytics, data mining, data visualization, data tools and infrastructure, and best practices to help organizations support data-driven decisions (Azeem, 2016).

There are cases in the primary sector where technologies like business intelligence are implemented to manage resources, for instance, AgtechApps, a company that uses a modular system that allows farmers to have digital control of data related to nutrition and inventory. With the data obtained, information, reports, and graphs can be generated automatically. Additionally, the system is easy to integrate and information is constantly available to be sent (Visualnacert, 2021).

Another case is Visualnacert, which implements multiple technologies for crop management and control. The corporation mentions that they have the "VISUAL Gestión Agronómica" tool, which is focused on monitoring pests, diseases, and treatment management (Agtechapps, 2021).

Case study

The Granja Acuícola de Occidente, located in the town of San Buenaventura, in Manzanillo, Colima, Mexico, the subject of this applied research, is dedicated to the reproduction, hormone treatment and cultivation of Stirling tilapia. This farm uses one or more ponds to raise tilapia fingerlings, and it takes up to seven months to reach the ideal size; some of the determinants for this growth are the water parameters, climate, frequency and amount of food, and even stressful situations for the species.

The farm's producer possesses a technology called C-TRA, a web system where the values associated with each production process, such as consumption of inputs, measurement of water quality parameters in ponds, product sales, among others, are recorded (Lerdo, 2021).

Despite this system, different problems occur during tilapia culture, expressively, low growth, decrease or increase of consumed feed, even diseases. Also, there is a lack of computational tools to process and exploit the recorded data, to give a possible basis for adversities and address them, or to corroborate good farming practices. Therefore, decision making is based mainly on the experience of the workers, causing inaccuracy in the causes of the events.

Additionally, productive efficiency is mostly calculated by applying a direct rule of three, where an approximate amount of fingerlings planted is related to the amount of species harvested, showing the lack of a continuous improvement process.

Literature review

There are multiple ways to perform an analysis using BI, for example, creating a model through all its stages, or using available software.

In the thesis work elaborated by (Michele, 2022), is presented the development of a quotation system for a real estate company using artificial intelligence and business intelligence. It is mentioned that the system is developed using the frontend tools none.js, vue.js, chart.js, making HTTP requests to an API. This API obtains the data from a PostgreSQL database and returns them to the same system. In addition to the above, they make use of MS Power BI, as support in the relationship of records and chart generation.

BI is present in several applications oriented to agriculture. (Bimonte et al., 2021) present an analysis of agro-diversity in French agricultural fields using an own business intelligence framework. To create this framework, they take records from the Farmland Biodiversity Observatory database, a science program that was designed for the country's agricultural activity. The retrieved records are transferred to a data warehouse using an Extraction-Transformation-Loading (ETL) tool. When cleansing the data, they mention that the analysis is through data science methods, where different levels are considered to target different types of users. The relational storage implements an OLAP tool in Mondrian server (Hyde, 2006), which relates dimensions such as agricultural systems, crops, fertilization, geographical location, field type, pesticide use, meadows, among others. With an OLAP tool in Saiku, a service that allows the creation of reports through draggable elements (OpenExpo, 2019), they display the information for clients or expert analysts, where each user has the facility to select the queries of personal interest, visualizing the results through bar charts, pie charts, among others.

Similarly, the work presented by (Dávalos, 2022) shows the development of an optimizing system for monitoring indicators in a brewing company, where business intelligence is used to generate reports and manage orders, stock, product volume, customers, among others. The development begins with the extraction of data stored in Microsoft Excel, where CSV files are produced. These records are retrieved through Pentaho software, which oversees the ETL process, cleaning those records that are not useful, and saving the rest in a database on PostgreSQL. Subsequently, they use the OLAP system in Saiku, which is responsible for managing the relationships between values, and producing information for the company.

Another way to generate information through data warehouses, is to use dimensional schemes that relate records in a simplified way, as in the case of (Ngo, Kechadi, 2020). In their study they use a constellation type scheme to relate data acquired from multiple sources. The schema has five fact tables, "field facts", "sales", "orders", "tests" and "management actions", and 22 dimensions. The values of the dimensions include field data, type of substrate, nutrients, quantity of sales, price per unit, products purchased, among others. Lastly, they reported three correlations, one between crops and soil type, where the factors analyzed are pH, phosphorus, potassium, and magnesium content; another correlation was between crops and the herbicides used, and the las one between crops and insecticides.

For the calculation of productive efficiency, several studies were located, which take as parameters the quality and quantity of feed supplied to the ponds, the duration of each crop, the size of the ponds, and the general knowledge of the farmers. According to the article presented by (Iliyasu, Mohamed, 2016), they use the Data Envelopment Analysis (DEA) methodology in one and two stages, mentioning its use due to its practicality, since it can be useful in economics, health, agriculture, transportation, education, and manufacturing.

December 2022, Vol.8 No.22 23-35

Conversely, the work published by (Yuan et al., 2020) makes use of Stochastic Frontier Analysis (SFA). This is predominantly used for the estimation of technical efficiency studies in the aquaculture sector, mentioning that SFA attributes all deviations from the production frontier to technical inefficiency, making it an inappropriate technique in some sectors, especially in agriculture, whose data collection process is sensitive to stochastic noise and other measurement errors.

Proposed solution

As described in the previous section, several technologies focus on the use of BI systems for the analysis of data from agricultural activities. Therefore, the importance of aquaculture as an economic and food activity should be highlighted.

Thus, based on the nature of the problem identified in the aquaculture farm, on the available data and on the free access tools that will allow a higher quality evaluation of the information, this work proposes to generate a system based on business intelligence, called the Intelligent System for Aquaculture Tilapia Management (SIGETA), which involves technologies for the design of data warehouses, ETL processing, statistical calculation and generation of graphs and tables.

The objective of this research is to present the modeling of SIGETA as a business intelligence system, managing the *Stirling* tilapia farming indicators, analyzing the production and parametric variables, the number of seeds and the harvest, along with the diseases of the species, supporting the understanding of the general situation of each pond, as well as the adequate management of personnel.

This proposal innovates in an incremental way, being that it uses the C-TRA database to acquire records, also, it reports efficiency indicators for each pond, contributing to the aquaculture sector. Once implemented, it will monitor the situation of each pond during tilapia farming, and the producer will identify the possible causes of decreases in production efficiency and can make decisions that contribute to maintain the expected efficiency or even increase it.

Article structure

The document is divided in seven sections: the first one corresponds to the introduction, with the statement of the problem, the literature review and the proposed solution; the second one, materials and equipment, mentions the technologies used for the construction of the system; continuing with the experimental methods, which specify the methodology used in the development of the research; the results contributed to the farm and to the system itself; the conclusions, emphasizing the fulfillment of the objectives and the benefits provided; finally, the acknowledgements and references sections are presented, the latter as support for the proposal.

Methodology

The case study is linked to the needs reported by the owner of the “Granja Acuícola de Occidente”, which is part of the aquaculture product system, recognized by state government agencies. The materials and methods used, due to their characteristics such as work flexibility, costs, support, among others, help to present a model of a competitive tool with respect to current proposals.

Materials

As an essential material is the database that allows C-TRA to operate, additionally, the same technologies used in this system are applied, with the purpose of give continuity and homogeneity to the code, only having as an added value the Pandas library, from Python, which will perform the statistical calculations.

Among the digital tools used is MySQL, the database management system. This tool manages all the records captured in C-TRA, in addition to the tables that make up the data warehouse and the queries necessary for system operation. For the visualization of SIGETA, HTML and CSS are used, both layout languages for structure and customization. Bootstrap, a CSS framework, is used to support usability. Regarding the manipulation of the events that occur in the application, Jquery, a JavaScript library, is used.

Additionally, the Python Pandas library is used, which is focused on statistical analysis, whose properties are the ease of loading data from various sources, such as CSV, JSON, HTML, or GBQ files, and the handling of large amounts of data is done efficiently compared to other tools such as Microsoft Excel, or Numpy, another statistical library for Python (McKinney, 2012).

Methods

Since this project is focused on the aquaculture area, and the properties and characteristics of tilapia culture are specified, the approach of this research is quantitative with descriptive character, where the productive efficiency is analyzed (Sampieri, 2014). In addition, an incremental product innovation in the technological area is detailed (Publishing, 2005).

The research takes as universe the culture ponds of the aquaculture farm, specifically the active ponds during the development of the modeling. To deepen the requirements of the system, semi-structured interviews were conducted with the owner of the farm; in addition, data was obtained by extracting records from the database that the producer has.

As a development methodology in the management of the software engineering process, the agile methodology Scrum was followed, given its characteristics of adaptation, iteration, speed, flexibility, and efficiency, designed to deliver significant value quickly in every project (SCRUMstudy, 2019). The Scrum workflow is shown in Figure 1.

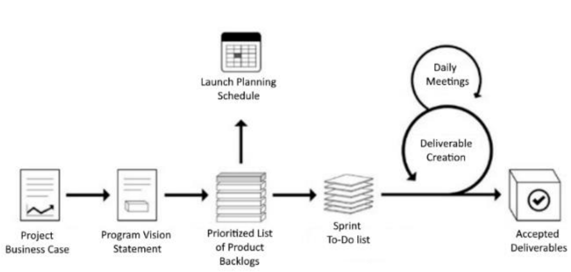

Figure 1 Scrum's project life cycle
Source: SBOK Guide 3rd edition

As can be seen, the business case is identified, the project vision is created and from this document, the prioritized list of pending issues is created containing the user stories that relate the functionalities required by the users. These are planned in sprints or iterations for delivery. Each sprint contains a certain number of user stories that become deliverable increments at the end of the sprint.

Along with Scrum, this research incorporated the Kimball methodology for data warehouse modeling, enabling fast record retrieval (Mendoza, 2020). The structuring of the data warehouse follows a bottom-up approach, where datamarts are created first, based on system requirements. The primary data sources are evaluated and processed using an ETL tool that captures and cleans the records, which are then loaded into a data warehouse, which relates the data in a fact table, as a support for the information to be generated.

The Kimball methodology aims to make data warehousing run quickly, providing analytical reporting capabilities for specific business processes, simplifying data management due to its non-standardized structure, which simplifies queries and analysis (Naeem, 2020).

Methodology application

The constant contact with the producer allowed to define the vision of the project, also considering the possible changes and risks during the development. The project vision included the epics and user stories that defined the scope of the project. Subsequently, eight sprints of two weeks each were planned. As each sprint progressed, results were obtained that were useful to the client, considering the priority of each task and the comments for the adequacy of the system, with each deliverable being reviewed and endorsed by the client. The process was repeated in a similar manner during the following sprints. As a result of the project vision, Figure 2 shows the conceptual model of SIGETA.

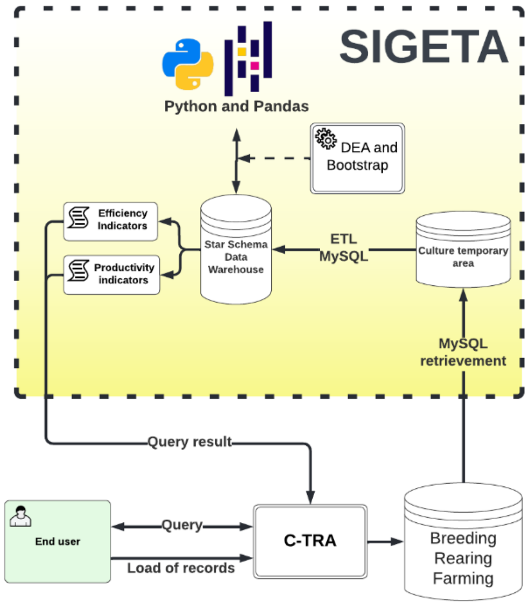

Figure 2 SIGETA's conceptual model
Source: Own elaboration

The process begins when the user (green box at the bottom left) loads the records using the C-TRA tool, which are stored in the system's database. The records contain the historical values of breeding, rearing, and farming periods. Subsequently, by means of MySQL queries, only the tables related to the culture are retrieved. These tables remain in a temporary or data preparation area; this database is considered the first action of SIGETA (yellow box at the top) and leads to ETL processing.

Employing the Pandas library, the ETL process is performed from the temporary area to the data warehouse, obtaining new tables, organized in a multidimensional model in the form of a star. This model has a fact table called efficiency, and six dimensions, which are *biometrics*, *parameters*, *outputs*, *inputs*, *feed*, and *medication*.

Once the warehouse is structured, relationships and analysis can be performed with the data, through preloaded queries in SIGETA. One of the analyses is the calculation of the productive efficiency, which first implements the DEA statistical method, followed by the bootstrap statistical method, which will give greater precision to the analysis. These calculations are carried out with Pandas, which takes from the data warehouse the relation of records that allow to generate the efficiency indicators.

In the same way, simple relations can be made to contemplate the situation of each pond, such as the relation between consumed feed and average growth, water parameters of the ponds against feed consumption, among others, resulting in production indicators. The results of the queries return graphs or tables for the end user to facilitate the understanding of the situation of each pond, thus supporting decision making.

Epics and user histories

A software life cycle contains processes, activities, and tasks applicable during the acquisition, provision, development, operation, maintenance or disposal of software systems, products and services, through the involvement of stakeholders with the objective of meeting customer needs (International Organization for Standardization, 2017).

Epic: 2	History N°: 1	Title	View creation of SIGETA
Priority	High	Risk	Low
Estimation	8h	Responsable	Sabino Tonatiuh González Rodríguez
Description	As user I want to have a SIGETA tab to see the situation of each pond		
Validation	Select from the C-TRA menu the SIGETA option. Display a query window To have a drop-down menu containing the ponds Have query options Contain a box where the graphs or tables will be displayed.		

Table 1 Example of a User History, describing a task and his validations
Source: Own elaboration

Given the planning of the eight sprints, in the progress of each one, useful results to the client were obtained, taking into account the priority of each story and the comments for the adequacy of the system, with each deliverable being reviewed and endorsed by the client. The process was repeated in a similar way during the following sprints.

Deliverables were based on epics, user stories and tasks. Epics are the stakeholder's requests, which give an overview of the project (Mendoza, 2020). In this research, five epics were obtained, visible in Table 1.

ISSN-2444-5002

No.	Epic
1	As a user I want to know the efficiency of each pond.
2	As a user I want graphs to be presented, detailing the calculations and analysis performed by the system.
3	As a user I want to compare efficiency between ponds.
4	As a user I want to save or print the generated tables or graphs.
5	As a user, I want the data to be updated every 15 days.

Table 2 Epics with SIGETA's scope
Source: Own elaboration

User Stories (UH) are numbered, they are assigned a priority value, a risk value, responsible parties and a general description are assign using the format "I like... want... for...", following the model proposed by Mike Cohn, a Scrum Triner co-founder of the Scrum Alliance, which answers three things, "What type of user is the user requesting?", "What is wanted?", and "Why is it wanted?" (Menzinsky et al., 2020). Table 2 exemplifies the user story about the pond situation.

The UH follows the aforementioned format, since it contains information regarding the epic it belongs to, the user story number, its name, the importance of its realization and complexity, as well as a numerical estimate to easily compare it with other user stories of the same epic, and finally, the name of the person responsible for its realization.

Therefore, this methodology helped to naturally adapt to changes and user requirements. Also, although Scrum is a structured methodology, its execution can be adapted to the needs of the Project (Drumond, 2020).

Efficiency calculation

DEA combined with bootstrap is used to estimate the productive efficiency of each pond. DEA is an alternative principle to generate information about an examined population, optimizing each individual observation with the objective of calculating a discrete piecewise frontier (Charnes, 1997). An example is presented in Figure 3.

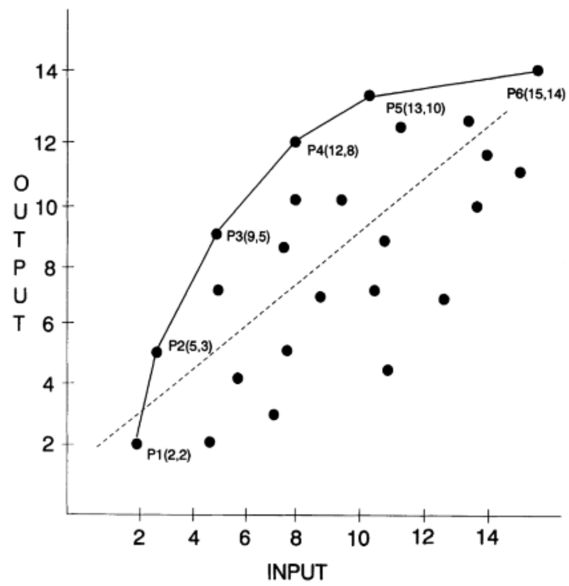

Figure 3 DEA example diagram
Source: Open Edition Journals

The graph shows the relationship of inputs on the abscissa axis versus outputs on the ordinate axis. An input is a resource or material that is used to generate a product, and an output is the amount or volume of profit generated by the input. For the case of the farm, there are several types of inputs and outputs, which depend directly on the desired analysis. For example, an input is the amount of feed consumed by the tilapia in a given time, and an output is the growth in grams obtained by the fish.

Each point on the graph represents the relationship between the number of resources used and how much benefit was obtained from it. Then, those with a closer approach to the output axis, adjacent to the input axis, are joined; the line indicates the cases that obtained the best result, and the rest of the points are considered as deficient cases.

Due to the alterations that can occur during the tilapia culture process, it is risky to assert that a pond is productive or not; the ideal is to repeat multiple times the production of tilapia in a pond, trying to replicate the conditions presented, being this virtually impossible.

Therefore, in addition to DEA, bootstrap is used, defined by (Efron, 1994) as a computer-based method that assigns measurements to approach statistical estimates.

This method takes the arrays of output values to create a new data set, or all the points of the DEA plot, this selection being randomized with the possibility of repetition, called sampling with replacement, preserving the same amount of values of the original set. Subsequently, the average is calculated and recorded in a histogram, exemplified in Figure 4.

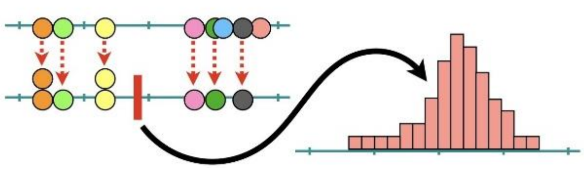

Figure 4 Bootstrap working example
Source: StatQuest with Josh Starmer

This process is repeated *n* times, where the number of iterations generated with bootstrap depends directly on the computational capacity of the equipment performing the calculations. It is recommended to do as many as considered reasonable, although increasing the number of iterations reduces the effects produced by random events or failures in original samples (Efron, 1979). According to (Efron, 1987), 50 iterations are sufficient to obtain up to a 95% confidence interval, being possible to have an estimate very close to the mean of the real efficiency value.

Results

The modeling of the SIGETA system was obtained, which includes the data warehouse on which it operates, as well as the views that compose it, supporting the aquaculture farm in the management of Stirling tilapia farming data.

SIGETA has web features, it considers the connection with the MySQL server database of the C-TRA system; using ETL, the data are extracted, transformed and loaded to the created warehouse, the data is arranged in a multidimensional star-shaped schema, and the records are processed with the Pandas library, where DEA and bootstrap methodologies are applied; finally, the production and efficiency indicators are generated, which the final user can visualize through graphs and tables.

System functionality

Users accessing SIGETA can select the interest to know by means of tabs, executing preloaded queries in the code, which speeds up the acquisition of results, like *Efficiency, General Information*, *Biometrics*, and *Relationship* as shown in Figure 5. Each tab has a section to select both the desired tank and the period of records, specifying the required start date and end date.

Figure 5 View *General info*, the main view of SIGETA
Source: Own elaboration

The tab *General Information* section shows graphs that reflect the values recorded directly in C-TRA associated with the farming stage, including physicochemical parameters. The B*iometry* tab indicates the growth values of the tilapia, the measurements taken, the average size and the average growth.

The *Relationship* tab, presented in Figure 6, performs relationships of multiple variables, for example, the amount of feed consumed versus growth obtained, or how applying a drug affects feed consumption and growth. This relationship performs a calculation based on DEA methodology. In the graphs, each point represents a date and what characteristics were present, identified by color. As an illustration, the red dots indicate July 1 as the date, but each chart shows how the input (feed, oxygen, temperature) is related to the output (growth, feed consumption).

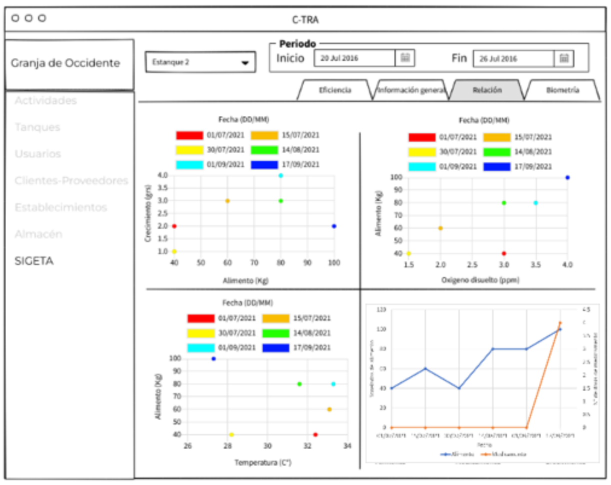

Figure 6 View Relation from SIGETA
Source: Own elaboration

The relationship of growth with respect to the amount of feed, shown in Figure 7, is plotted in the upper left panel of the page, and indicates as input the amount of feed supplied to the pond, measured in tons, and as output the growth of the tilapia, expressed in grams. This growth is the comparison of the previous weight versus the current weight.

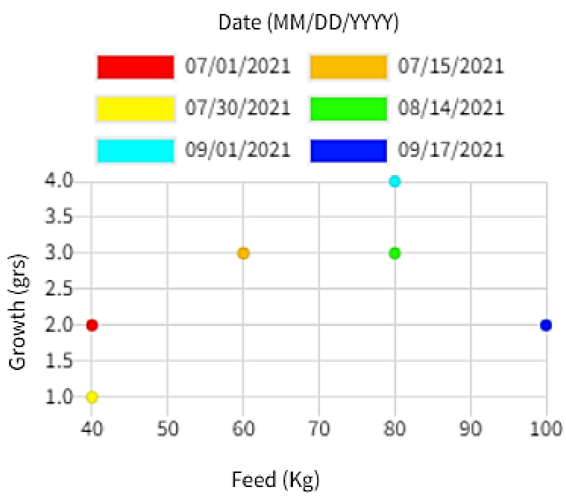

Figure 7 Detail of the graph showing the relationship between feed consumption and growth
Source: Own elaboration

The upper right panel indicates feed consumption versus the amount of dissolved oxygen in the water, visible in Figure 8, taking oxygenation as input and feed consumed as output. Both values are averages of the lapses between biometrics.

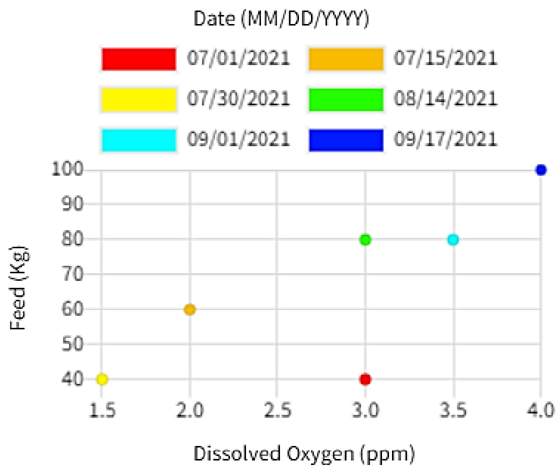

Figure 8 Detail of the relationship between dissolved oxygen and feed consumption
Source: Own elaboration

The lower left panel, seen in Figure 9, reflects the amount of feed supplied (horizontal axis) with respect to the temperature in degrees Celsius of the pond (vertical axis).

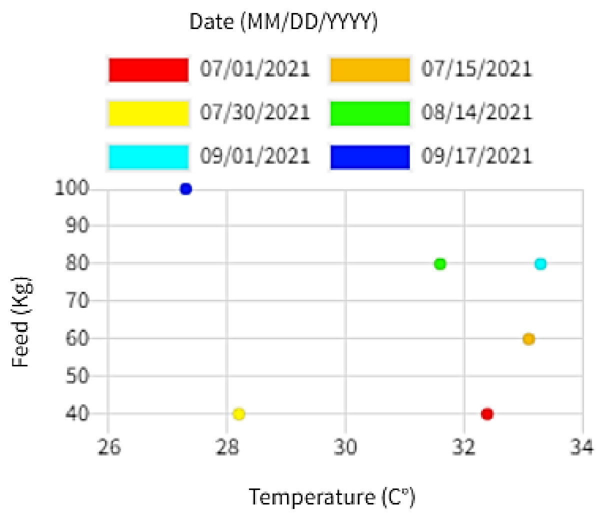

Figure 9 Detail of the graph relating pond temperature and feed consumption
Source: Own elaboration

The relationship of feed intake to drug supply, Figure 10, shows how drug doses can affect feed consumption.

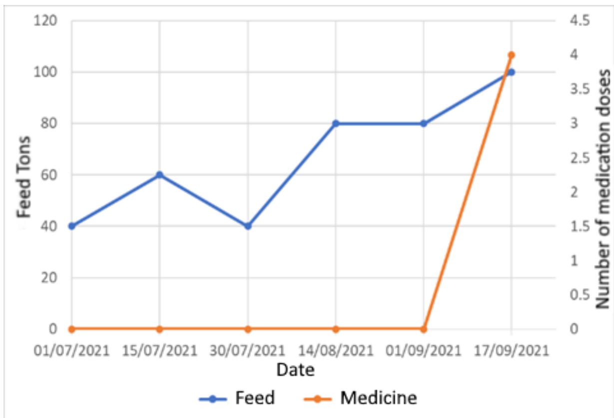

Figure 10 Detail of the food-medicine relationship graph
Source: Own elaboration

It should be noted that the use of the DEA generates an approximation of the efficiency of each pond between measurement periods, but there may be uncontrollable factors that affect tilapia growth, such as stress due to the presence of predators, cloud cover that darkens the ponds, excess nitrite and nitrate from fecal matter, among others.

To increase the precision of the productive efficiency of each pond, the bootstrap methodology was used. The output values provided by DEA are used to re-sample with substitution, creating new arrays of values.

In this case, the generation of the bootstrap indicator is related to the dates selected by the user, since they determine the number of samples taken for the arrays. Following the procedure, the re-sampling is taken randomly, with the possibility of repetition of values. Given the computational capacity of the producer's equipment, 50 iterations are performed, as observed in 2.4, in order to create highly reliable indicators, without sacrificing processing speed and consumption of computational resources.

The average indicator obtained by bootstrap is the one used in the graphs and tables of the efficiency tab, shown in Figure 11, reflecting a quasi-rational value of the pond's productive efficiency (lower right section of the Figure).

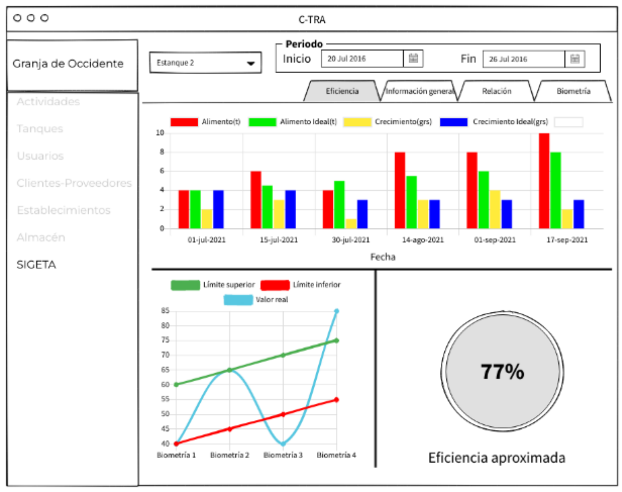

Figure 11 Efficiency tab view from SIGETA
Source: Own elaboration

Of these graphs, the one located in the top panel, shown in Figure 12, compares the feed consumed against an ideal consumption, based on the farm's historical record. It also compares the given growth against the ideal, taking the same parameters as the feed.

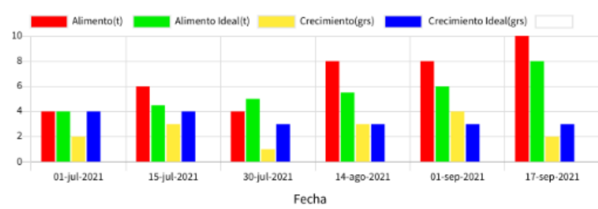

Figure 12 Comparison of indicators. Real versus ideal values
Source: Own elaboration

Likewise, the figure 13 shows a graph that denotes an upper and lower limit, with the upper limit being the ideal growth of the tilapia, and the lower limit being that growth considered critical, where any value close to this indicates special attention. Likewise, the graph shows a line of the value taken by the biometries.

ISSN-2444-5002

December 2022, Vol.8 No.22 23-35

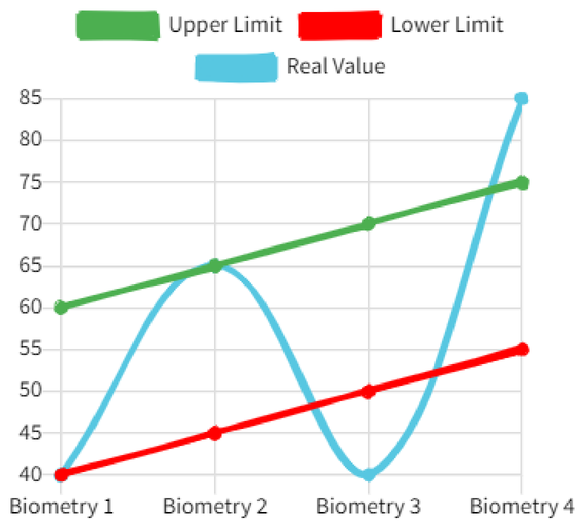

Figure 13 Detailed graph of maximum expected growth vs. minimum desired growth
Source: Own elaboration

Data Warehouse

The logical part of data management refers to the records in the data warehouse, which are maintained locally. The data warehouse is based on the Kimball methodology, using a star-shaped model, shown in Figure 14, which includes the fact table "*efficiency*", along with the dimensions *parameters, income, medication, feed, outflows* and *biometrics*. This scheme facilitates the collection of data associated with production and efficiency indicators, in relation to the amount of tilapia harvested, as well as the behavior in relation to feeding behavior and physicochemical parameters.

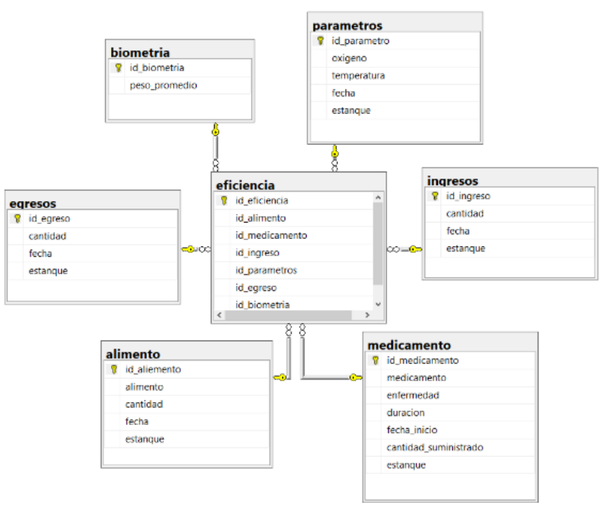

Figure 14 Data Warehouse of SIGETA in star-shaped schema
Source: Own elaboration

The design of SIGETA, as well as the data warehouse, meet the requirements of the farm producer, as the epics that defined the scope of the system are addressed.

Discussion

The results show the modeling of the SIGETA system; the visual part for the user, as well as the functioning of the data warehouse on which it operates. It can be appreciated the variety of data that can be analyzed and related, and how this reflect behaviors in response to certain stimuli during tilapia culture.

This modeling proposes a solution to the problems of the aquaculture farm owner and the C-TRA system, allowing the generation of information through graphs that support decision making.

Regarding related work, SIGETA combines statistical calculations of efficiency with management, using computational technologies. In addition, the use of this system will clarify uncertainties in the face of complications, substantiating the empirical knowledge of the workers.

Acknowledgments

To the Consejo Nacional de Ciencia y Tecnología (CONACYT) and the Instituto Nacional de México campus Instituto Tecnológico de Colima, for allowing me to continue my academic training and contribute to technological progress in the state of Colima. Also, to the Granja Acuícola de Occidente for facilitating its installations and processes for this case study.

Financing

This work has been funded by CONACYT, 803437.

Conclusion

This applied research presents SIGETA, whose contribution to the field of knowledge is the modeling of the data warehouse and the system views, being this a process innovation from the analysis of information from the records that follow up the tilapia farming; and a product innovation, by providing a new system for the company.

The views work for the analysis and management of data produced during the tilapia aquaculture farming stage, taking advantage of the fact that the producer has records of tilapia culture monitoring through C-TRA. Thanks to the data, production and efficiency indicators can be generated to support decision making.

The impact of the system is to provide an opportunity for the producer to know the behavior of each pond, and to monitor productive variables, such as feed, temperature, oxygenation, growth, and use of medications. The result of permanent monitoring promotes the management of the resources invested during cultivation.

SIGETA represents direct benefits for the producer, since at any moment of the farming he can relate the productive variables, looking for situations that can cause affectations and act on the basis of what has been observed. This type of system can be implemented in other aquaculture activities of the primary sector, following the same support principles.

On the other hand, we have a low-cost system that makes use of open-source software, allowing the scalability of the system and facilitating the maintenance of the code.

In terms of software development management methodology, the effectiveness of Scrum with Kimball integration, expedites the work and deliveries, anticipates risks and adapts to customer needs.

As future work, the implementation of machine learning systems is proposed, as well as the taking of external factors, such as the presence of natural phenomena or data from external sources, to generate a prediction of tilapia behavior.

References

Agtechapps. (2021). Control y monitoreo de actividades agrícolas. https://www.agtechapps.com/activity-app/

Azeem H. (2016). Business Intelligence y análisis de datos: obtención de valor de su almacén de datos. Astera. https://www.astera.com/es/type/blog/data-warehouse-and-business-intelligence/

Bimonte, S., Billaud, O., Fontaine, B., Martin, T., Flouvat, F., Hassan, A., Rouillier, N., & Sautot, L. (2021). Collect and analysis of agro-biodiversity data in a participative context: A business intelligence framework. Ecological Informatics 61. https://doi.org/10.1016/j.ecoinf.2021.101231

Charnes, A., Cooper, W., Lewin, A. Y., & Seiford, L. M. (1997). Data envelopment analysis theory, methodology and applications. Journal of the Operational Research society.

CONAPESCA. (2021). La Conapesca promueve la producción y consumo de tilapia en el país. https://www.gob.mx/agricultura/yucatan/articulos/la-conapesca-promueve-la-produccion-y-consumo-de-tilapia-en-el-pais

Dávalos, M. (2022). Desarrollo de un sistema de Business Intelligence que optimice el monitoreo de indicadores de productividad en la empresa “DINADEC Centro de Distribución Latacunga” de la provincia de Cotopaxi. [Tesis de licenciatura]. Universidad de las fuerzas armadas.

Drumond, C. (2022). Scrum Aprende a utilizar scrum con lo mejor de él. https://www.atlassian.com/es/agile/scrum#:~:text=Scrum%20est%C3%A1%20estructurado%20para%20ayudar,pueda%20aprender%20y%20mejorar%20constantemente.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics, 7(1). https://doi.org/10.1214/aos/1176344552

Efron, B. (1987) Better Bootstrap Confidence Intervals, Journal of the American Statistical Association, 82:397, 171-185, DOI: 10.1080/01621459.1987.10478410

Efron, B. & Tibshirani, R. J. (1994). An Introduction to the Bootstrap. Taylor & Francis.

FAO. (2020) El Estado Mundial de la Pesca y la Acuicultura 2020. La sostenibilidad en acción. Roma. https://www.fao.org/documents/card/en/c/ca9229es

FAO. La acuicultura y las pesquerías basadas en el cultivo. https://www.fao.org/3/y5751s/y5751s08.htm#fn5

Huaman, M. (2022). Sistema de cotización con Inteligencia Artificial e Inteligencia de negocios en la empresa Buenavista Proyectos Inmobiliarios SAC. [Tesis de licenciatura]. Facultad de ingeniería y arquitectura escuela profesional de ingeniería de sistemas.

Hyde J. (2006). Mondrian Documentation. https://mondrian.pentaho.com/documentation/olap.php

Iliyasu, A. & Mohamed, Z. A. (2016). Evaluating contextual factors affecting the technical efficiency of freshwater pond culture systems in Peninsular Malaysia: A two-stage DEA approach. Aquaculture Reports, 3, 12-17. https://doi.org/10.1016/j.aqrep.2015.11.002

INAES. (2018). Acuicultura, historia y actualidad en México. https://www.gob.mx/inaes/articulos/acuicultura-historia-y-actualidad-en-mexico#:~:text=La%20acuicultura%20se%20enfoca%20en,remontan%20al%20a%C3%B1o%203800%20a.c.&text=La%20acuicultura%20o%20acuacultura%20es,especies%20acu%C3%A1ticas%20vegetales%20y%20animales

Ipac Acuicultura. (2021). Las nuevas tecnologías empleadas en acuicultura son confusas o desconocidas para los consumidores europeos. http://www.ipacuicultura.com/noticias/en_portada/79801/las_nuevas_tecnologias_empleadas_en_acuicultura_son_confusas_o_desconocidas_para_los_consumidores_europeos.html

International Organization for Standardization (2017). Systems and software engineering — Software life cycle processes (ISO/IEC/IEEE 12207:2017) ISO/IEC/IEEE 12207:2017

Lerdo H. (2021). Sistema Web para la Trazabilidad de la Producción y Comercialización de Tilapia en Granjas Acuícolas [Tesis de maestría]. Tecnológico Nacional de México.

McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. "O'Reilly Media, Inc.". ISBN-10: 1491957662

Mendoza A. (2020). Metodologías de Data Warehouse. Gravitar. https://gravitar.biz/datawarehouse/metodologias-data-warehouse/#:~:text=Un%20Data%20Warehouse%20es%20un,consultas%20y%20toma%20de%20decisiones.

Menzinsky, A., López, G., Palacio, J., Sobrino, M., Alvarez, R., & Rivas, V. (2020). Historias de Usuario: Ingeniería de Requisitos Ágil. Scrum Manager. https://www.scrummanager.com/files/scrum_manager_historias_usuario.pdf

Naeem T. (2020). Conceptos de Data Warehouse: enfoque de Kimball vs. Inmon Astera. https://www.astera.com/es/type/blog/data-warehouse-concepts/

Ngo, V. M., & Kechadi, M. T. (2020, January). Crop knowledge discovery based on agricultural big data integration. In Proceedings of the 4th International Conference on Machine Learning and Soft Computing (pp. 46-50).

OpenExpo Europe. (2019). Saiku y ctools para mejorar el análisis de datos. https://openexpoeurope.com/es/saiku-y-ctools-para-mejorar-el-analisis-de-datos/

Publishing, O. (2005). The Measurement of Scientific and Technological Activities Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data, 3rd Edition (3rd ed.). Org. for Economic Cooperation & Development.

Sampieri, R., Collado, C., & Lucio, P. (2014). Metodologia de la investigación. 6ª ediction.

SCRUMstudy. (2019). A Guide to the Scrum Body of Knowledge (SBOK Guide). VMEdu Inc.

Visualnacert. (2021). VISUAL Gestión Agronómica. https://visualnacert.com/visual-gestion-agronomica/

Yuan, Y., Yuan, Y., Dai, Y., Zhang, Z., Gong, Y. & Yuan, Y. (2019). Technical efficiency of different farm sizes for tilapia farming in China. Aquaculture Research, 51(1), 307-315. https://doi.org/10.1111/are.14376.