Use State-Of-The-Art Data Mining Technologies To Improve Feeding Strategies And Predict Feed Conversion Rate - The Aquasmart Solution

Konstantinos Seferis¹, Gerasimos S. Antzoulatos¹, Gary McManus² and Tom Flynn³

¹Integrated Information Systems (I2S), Greece
²Waterford Institute Of Technology (TSSG) Coordinator, Ireland
³Q-Validus Limited, Ireland
What is the need?

How can I maximize the growth rate of my livestock?

How can I minimize the production costs?

How can I reduce the impact to the environment?

Is my performance worse or better than the others?

Daily main concerns of an aquaculture producer
What is the need?

Unfortunately, answering these questions is not that simple

Parameters to consider:

- environmental,
- feed types & composition,
- feeding rates & practices,
- net changes,
- production management strategies,
- the people involved
- and many others…

Drowning in data, but…

difficult (almost impossible) to:

- Understand,
- quantify the interaction between the parameters manually,
- detect trends, regularities, patterns,
- identify rules,
- interpret

starving for the knowledge that will help to reduce costs, improve operational efficiency and increase performance!!!
AQUASMART can be the answer to these needs!!!
The solution is aquaSmart...

aquaSmart is a European Big Data project for the Aquaculture sector, funded by the European Commission’s Horizon 2020 Program

**Objectives**

- Enhances **innovation capacity** within the aquaculture sector
  - by helping companies to transform captured **data** into **knowledge**
  - and use this **knowledge** to make **decisions** to improve performance

- Promotes **best practices** for aquaculture production management

- Develops a **training program** and **training activities** for the people of aquaculture
Mission Statement

AQUASMAR supports:

- Lower production costs
- Higher profitability
- Improved operational efficiency
- Carry out business in a sustainable, environmentally friendly way

from 45 million tons in 2014 to 85 million by 2030

define best practices for the industry

Sustainable Growth
Innovation

- Completely novel and unique approach
- Provide highly accurate **descriptive** and **predictive** models to understand not only *what has happened*, but *what is most likely going to happen next*
- Our team consists of experts from academia and experienced personnel from Aquaculture companies

Adding **extra Business Value** and **Competitive Advantages**
AQUASMART Approach

Analyze → Model
Business Questions

• Some of the Business Questions that are answered by AQUASMART:
  ✓ Evaluation of fish population performance
  ✓ Evaluation of feed performance
  ✓ Evaluation of feeding
  ✓ Evaluation of fry quality
  ✓ Evaluation of the influence of the environment
  ✓ Estimation of fish number and average weight
Lets get Practical

How can use AQUASMART?
Aquasmart Analytics

1. Upload dataset
2. Define a business question
3. Preprocess the dataset by:
   a. regulating the filters to remove extreme values or “noisy” entries
   b. subsetting the dataset choosing the appropriate labels of variables to analysis (ex. choose only summer months)
4. Execute the Descriptive Statistics Analysis so as to gain a good insight of the data (detect trends, correlations etc)
5. Create Machine Learning & Data Mining models to discover the underlying “hidden” knowledge of the data, which is related to the research hypothesis
Scenario

- A farm manager wants to explore production data using Descriptive Statistics tools in order to improve feeding.

Steps

1. Collect the appropriate data for analysis and prepare the dataset
2. Upload the dataset to aquaSmart
3. Preprocess the data
Test Case Scenario - Preprocessing

- Is very important in order to make a good analysis and develop models that work
- It is done by viewing the attributes of the population using the tools provided by aquaSmart
- The goal is to remove extreme values (wrong data or extreme fish populations).
- One way to do this is to produce density plots
Test Case Scenario - Preprocessing

- Regulate the filters:
  - $0 \leq \text{SGR} \leq 3$
  - $0 \leq \text{Econ FCR Period} \leq 4$
  - $0 \leq \text{SFR Period} \leq 4$
  - $\text{Avg Temp} \geq 15$
  - $0 \leq \text{Mortality}\% \leq 3$
  - $10 \leq \text{Days Between Sampling} \leq 60$
General Steps (*cont*):

6. Use **Descriptive Statistics** and **Visualization tools** so as to perform exploratory analysis on the data to accomplish several tasks, such as:

- View data distributions
- Identify skewed predictors (*skewness is a measure of asymmetry of a data distribution*)
- Identify outliers
- Other deviations from normality
Test Case Scenario – Descriptive Stats
Test Case Scenario - Descriptive Stats

- feed category 3 exhibits **higher SGR** values compared with the other feed categories.
- feed category 3 exhibits **lower Econ FCR Period**
- feed categories 2, 4 and 5 present **similar** behavior in terms of the distribution of SGR, FCR and SFR values.
- feed category 1 have much **higher FCR and SFR values** compared with feed categories 2, 4 and 5.
Why the behavior of Feed Category 3 is different?
**Test Case Scenario – Descriptive Stats**

**Why Feed Category 1 has FCR higher than others, but similar SGR?**

Feed Category 1 has higher FCR mean values in all of the Weight Categories (Start & End) comparing with the other feed categories.
The descriptive statistics showed that the feed type has a significant influence on the period FCR. The next step is to make a model that is able to predict the FCR of a period by taking as an input the following variables:

- Thermal profile of a region
- Period between two successive samplings (in days)
- Category of the feed, the Specific Feed Ratio (SFR) in the specific period of time
- Period (month) of sampling
Approach: He should create a Data Mining model based on historical periodic data

He should choose the variable to estimate of a drop-down list

He should choose the production factors (called predictors)

He should choose one of the available Machine Learning/Data Mining methods and launch the process.
Evaluate the trained model

- **Root Mean Square Error (RMSE):** *ideally tends to 0*
- **R-Squared:** *ideally tends to 1*
- In our case:
  - $\text{RMSE} = 0.067$ and
  - $\text{R-Squared} = 0.856$
Test Case Scenario – Machine Learning

Predict with the Model: now he can use the trained model and predict the FCR value in a new population
Test Case Conclusions

AQUASMART data analytics platform helps aquaculture companies to:

- **Visually explore** their data and **draw some primary conclusions**
- Detect fish populations with **abnormal behavior (outliers)**
- Create and train **reliable Machine Learning models** in order to
  - predict production KPIs for new populations
  - determine which populations exhibit expected (unexpected) behavior
  - evaluate the influence of the production factors
- **Reduce costs** through **reduction of food waste** without reduction in the fish growth. This helps companies to **optimize farm performance** and get the maximum return on their invested time, money and other resources.
- **Eliminate** or **reduce** the **overfeeding** and/or **underfeeding**, which have an undesirable impact on the growth and health of the fish as well as on the environment.
Conclusions – Future Work

- aquaSmart assists aquaculture companies to convert data into useful knowledge.

- We are committed to improving the fish production through the science of Statistics, Data Mining and Machine Learning. And all that in the context of adding extra business value into your corporate data

- The project goes much further: through the availability of multi-lingual Open Data, companies will be able to compare their result of their analysis with the ones that are stored into the cloud. This will generate a knowledge base that will be of maximum usefulness for all the stakeholders of the aquaculture sector

- The tool will be complemented by a comprehensive training program
Thank you for your attention!!!