UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

FERNANDA CARVALHO PINHEIRO

IN-LINE REAL-TIME WATER ACTIVITY PREDICTION BASED ON SOFT SENSORS: a case study in a pet food industry

Recife
2019

FERNANDA CARVALHO PINHEIRO

IN-LINE REAL-TIME WATER ACTIVITY PREDICTION BASED ON SOFT SENSORS: a case study in a pet food industry

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Pernambuco, como requesito parcial para a obtenção do título de Mestre em Engenharia de Produção

Área de concentração: Pesquisa Operacional.

Orientadora: Profª Dra. Isis Didier Lins.

Recife

2019

Catalogação na fonte
Bibliotecária Valdicea Alves, CRB-4 / 1260

P654l Pinheiro, Fernanda Carvalho.
In-line real-time water activity prediction based on soft sensors**:** a case study in a pet food industry - 2019.
61 folhas, Il. e Tabs.

Orientadora: Profª Dra. Isis Didier Lins.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia de Produção, 2019.
Inclui: Referências.
Texto Inglês.

1. Engenharia de Produção. 2. Aprendizagem de máquina.
3. Soft sensor. 4. Atividade de água. 5. Medição em tempo real.
6. Ração animal. I. Lins, Isis Didier. (Orientadora). II. Título.

UFPE

658.5 CDD (22. ed.) BCTG/2019-191

FERNANDA CARVALHO PINHEIRO

IN-LINE REAL-TIME WATER ACTIVITY PREDICTION BASED ON SOFT SENSORS: a case study in a pet food industry

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Pernambuco, como requesito parcial para a obtenção do título de Mestre em Engenharia de Produção.

Aprovada em: 26/02/2019.

BANCA EXAMINADORA

_____

Profª. Dra. Isis Didier Lins (Orientadora)

Universidade Federal de Pernambuco

_____

Prof. Dr. Márcio José das Chagas Moura (Examinador Interno)

Universidade Federal de Pernambuco

_____

Profª. Dra. Andrelina Maria Pinheiro Santos (Examinadora Externa)

Universidade Federal de Pernambuco

ACKNOWLEDGEMENTS

I wish to express my gratitude first to God, for the opportunity that I was given to carry out the Master's Degree. Also for having placed in my path professional people to whom I have respect and admiration.

Thanks to my parents, Rejane Carvalho and Fernando Pinheiro, as well as my brothers who supported me in my choices, and the other people who gave me support so that I could finish this stage. I thank my Professor Isis Didier Lins, Professor Márcio Chagas, and PhD student Caio Souto, who gave all the support to develop this work.

I would like to extend my thanks for Professor Andrelina Pinheiro, for always being willing to heal my doubts and contribute to my development. I thank my friends, old and new, that I will always carry in my thoughts and heart, together with all shared memories and learning.

I also thank the PPGEP - UFPE program for the opportunity to study the Academic Master's Degree and for all the teachings obtained during these two years.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil (CAPES) - Finance Code 001.

ABSTRACT

Water activity is considered an important parameter of quality, which represents the amount of water available for biochemical and chemical reactions, which enables the growth of microorganisms. It also contributes to food sensory characteristics, like texture and flavor. Water activity measurements can be performed by different equipment and methodologies. Using traditional equipment, a sample need to be collected and placed in a closed chamber to attain equilibrium. This procedure does not provide real-time values and may be associated to increased costs due to production out of specifications if the observed water activity is not in the desired range. The goal of this work is to fill the gap to monitor water activity value in-line in order to increase the response time, control process and make any nonconformity in the production readily verified. As a consequence, a reduction of costs is expected. This work proposes a methodology based on soft sensors to predict and control water activity from moisture content values. As identified in the literature, there is a relationship between water activity and moisture. However, it is not simple to characterize it once it is unique for each food. For that, different machine learning (ML) techniques (SVM, LS-SVM, MLP, GPR and LR) are adopted to map and learn this relationship The Statistical Process Control (SPC) methodology was also proposed as tool to monitor the soft sensor accuracy and to indicate when the underlying model may be retrained. The proposed soft sensors were applied to the case of a pet food industry. The results for all ML models were compared in order to guide the selection of which one would be adopted. All models had good performance, but GPR presented the best balance between model accuracy and training time.

Keywords: Machine learning. Soft sensor. Water activity. Real-time measurement. Pet food.

RESUMO

A atividade de água é considerada um importante parâmetro da qualidade, que representa a quantidade de água disponível para ocorrerem reações bioquímicas e químicas, viabilizando o crescimento de micro-organismos. A atividade de água também influencia as características sensoriais dos alimentos, como textura e sabor. A sua medição pode ser feita por meio de diferentes equipamentos e metodologias. Nos equipamentos tradicionais, uma amostra do alimento precisa ser coletada e colocada em uma câmara fechada até atingir o equilíbrio. Este procedimento não fornece valores em tempo real e pode estar associado ao aumento de custos devido a produtos fora de especificação, se o valor de atividade de água não estiver na faixa desejada. O objetivo deste trabalho foi preencher esta lacuna no monitoramento de atividade de água em linha, em função de reduzir o tempo de resposta, aumentar o controle do processo e fazer com que qualquer não conformidade na produção seja verificada rapidamente. Como consequência, uma redução de custos é esperada. Este trabalho propõe o uso da metodologia baseada em *soft sensors* para prever e controlar a atividade de água a partir de valores de umidade. Como identificado na literatura, existe uma relação entre atividade de água e umidade. Contudo, não é algo simples de caracterizar, uma vez que é única para cada alimento. Assim, diferentes algoritmos de *machine learning* (SVM, LS-SVM, MLP, GPR e LR) são adotados no mapeamento e aprendizado desta relação. A metodologia de controle estatístico de processos também foi adotada para realizar o monitoramento da precisão do *soft sensor*, bem como para apontar caso o modelo utilizado no *soft sensor* precise ser retreinado. Os resultados para todos os modelos de *machine learning* foram comparados a fim de guiar a seleção de qual seria adotado. Todos os modelos apresentaram performance satisfatória, porém o modelo GPR apresentou o melhor balanço entre tempo de treinamento e exatidão.

Palavras-chave: Aprendizagem de máquina. *Soft sensor*. Atividade de água. Medição em tempo real. Ração animal.

LIST OF FIGURES

TABLE LIST

SUMMARY

1 INTRODUCTION

In this chapter, the opening remarks, the justification, the objectives and the structure of this dissertation are presented.

1.1 OPENING REMARKS

The proliferation of microorganisms is one of the main factors that affects food perishability and shelf-life. For this reason, equipment and methodologies of analysis have been developed to guarantee food's quality, safety and traceability, in accordance with legislation and costumer's demand (CIFUENTES, 2012).

The microbial population behavior in food (growth, survival and death), is determined by its properties, like water activity (a_w) and pH, also by the environment conditions where the food is stored (temperature and relative humidity). These properties should then monitored to guarantee the final product's quality and durability (MCMEEKIN *et al.*, 1997).

Specifically, a_w is one of the main quality variables that impacts on shelf-life, texture, flavor, color and aroma of food, and it is also an indicator of product stability and microbial safety. Due to the relation of a_w with these food characteristics, U.S. Food and Drug Administration uses a_w as one of the references to evaluate food quality (FRANKS, 1991; LEWICKI, 2004 FDA, 2015).

In food science, a_w is useful as a measure of the potential reactivity of water molecules with solutes. In fact, it is very important to know the characteristics of a solution involving various components. However, in practice, it is only possible to determine easily the activity of binary solutions. The more complex the solution is, the more difficult it will be to measure a_w, once there are more interactions between different components and modifications of water behavior (CAZIER *et al.*, 2007).

Due to its importance, there are different methodologies and equipment to make its measurement. Nowadays this variable is measured in laboratory. Even though some of them can provide a result in about five minutes, there is a risk of sample's contamination, once it is necessary to collect the sample in the production line and take it to a laboratory. The process of taking samples and measuring a_w demands time. Without real-time measuring, product quality can get out of specification without being noticed and it would be perceived after the laboratorial analysis. When it happens, the output must be reprocessed, which means loss of time and raw-material, necessity of rework and cost increasing (NIELSEN *et al.*, 2015).

There are automated sensors to obtain other quality variables measurement, to control them in the process, and avoid some of the above-mentioned problems. Depending on the kind of sensor chosen, some advantages and disadvantages can be pointed out: In "on-line" case, an automated system is used to extract a sample that is sent through an specific duct to the sensor, which is not part of the main production line, for measurement purposes. The disadvantage of a sensor path separate from the main process is overcome with the development of "in-line" analyses, which is in direct contact with the process line, interacting directly with the sample. It avoids sampling steps, presents more representative measures, and the measurement occurs in shorter time. But, the direct interaction with the process, can cause sensors' wear and obstruction. Besides, it needs to be able to operate under extreme conditions of temperature and pressure. Another class of analysers is called "non-invasive", as the analyzer does not come into direct contact with the sample, it does not cause change in sample's composition or contamination. An analyzer can be qualified as 'real-time' if the time from two or more measurements and its processing by the central system does not exceed the hold-up time, which related to the time interval when a particle get in and out a system.(TREVISAN & POPPI, 2006).

A possible solution to obtain real-time a_w measurements and the necessity of a different equipment would be a soft sensor, which is a predictive model that uses available in-line/on-line/non-invasive sensor measurements for estimation of variables of product's quality that cannot be easily measured. In order to obtain these variables in real-time, computational intelligence methodologies could be used to build sensors to infer the value of the target variables from other real-time measured process variables. The basis for building such intelligent sensors is that the values of the target variable have a functional relationship with other process variables that can be measured in real-time (KADLEC *et al.*, 2009; SOUZA *et al.*, 2016).

The profitability in the implementation of sensors in-line occurs due to the increase in the optimization and the control of the process. However, it will always be necessary to have an analytical laboratory to carry out measurements, calibrations and maintenance of the sensors distributed throughout the process (TREVISAN & POPPI, 2006).

Given that, this work proposed a methodology for construction and maintenance of soft sensors to obtain a_w from other process variables (e.g., moisture content) that can be easily measured in-line using a sensor equipment, such as a humidity sensor, which can provide the value of food moisture content, a Near Infrared Spectrometer (NIR), that can provide in addition to the moisture content, other parameters such as crude fiber, crude

protein, ethereal extract, and so on. In the case of moisture content, literature have reported that it has a non-linear relationship with water activity, known as moisture sorption isotherms.

Modeling a sorption isotherm is not trivial, mainly for complex food with many ingredients an diverse composition, and the interaction between these components and the water in the food can interfere in the water activity value. What turns it unique for each food. Alternatively, in this work, the computational intelligence methodologies, to build the proposed soft sensor, were based on machine learning techniques: Support Vector Machine (SVM), Least Square Support Vector Machine (LS-SVM), Multi-Layer Perceptron (MLP), Linear Regression (LR) and Gaussian Process (GP). They were used as inference models to map the measured input values into an a_w point estimate. These techniques were applied to data from a pet food industry, using moisture content and a_w values, whose performances were compared. These algorithms were also combined to bootstrap methods to compute a_w interval estimates, which hold information about the precision of the obtained values. After building the soft sensor, a methodology for maintenance using the concepts of statistical process control (SPC) was proposed, since that, as other equipment, it may need maintenance and calibration. If results are different from the expected values, it is an indication that the soft sensor may be calibrated, that is, the underlying machine learning model should be retrained. Through this monitoring it is possible to verify whether the soft sensor is requiring a retraining or not, to ensure its proper operation.

The resulting soft sensors are expected to overcome the main disadvantages of the portable instrument- and laboratory-based a_w measurements: time delay, sample collection and possible contamination, direct operator participation, loss of an increased amount of resources if a_w is identified outside the required standards.

1.2 JUSTIFICATION

The investigation of a solution to make in-line a_w prediction for compound food (e.g., pet food) is worth of investigation, given that their corresponding a_w is often difficult to be determined due to the intricate interaction between water and the different food components that can change how much bond or free water is in the food. This interferes the value of a_w and also food quality characteristics.

As a_w has direct relation with moisture content, ML techniques were used to relate them, given that they can successfully deal with non-linear associations without requiring previous knowledge about the exact mapping between input (moisture content and possibly other variables) and output (a_w). Additionally, ML techniques have presented satisfactory

results when applied to model the relationship between process variables in other types of industry, and also when used to build soft sensors as pointed by Kadlec *et al.* (2009) and Yan *et al.* (2004).

A soft sensor to compute a_w in-line improves the measurement process as it becomes faster and does not require direct operator action in short time intervals. As a consequence the a_w soft sensor permits to: (i) rapidly identify and solve problems related to this food quality variable, which reduces loss of resources and the associated increased costs when compared to traditional manners of measurement; (ii) avoid sample gathering that could possibly be contaminated; (iii) reduce the operator participation in the quality measurement process.

With respect to (iii), personnel are still required to assess a_w using portable instrument or laboratory-based for maintenance purposes of the soft sensor. However, the frequency of such an activity can be significantly reduced and the professionals can be assigned to perform other tasks deemed labor intensive.

1.3 OBJECTIVE

The general and specific objectives are presented below.

1.3.1 General Objective

The general objective of this dissertation was to develop a methodology of construction and maintenance of a soft sensor to make in-line real-time water activity prediction, allowing its monitoring, as well as product maintenance under the desired specification. The proposed methodology was applied to a case study involving a pet food production process.

1.3.2 Specific Objectives:

In order to achieve the general objective, some specific goals are defined:

- Investigate current measurement techniques for a_w determination and their respective advantages and shortcomings;
- Investigate parameters that can be automatically measured by sensors and have relation with a_w or can influence its behavior in foods.
- Investigate machine learning techniques.
- Propose a methodology to construct a soft sensor based on ML techniques to measure a_w, taking into account the procedures to obtain, treat and analyze data.
- Use bootstrap methods to obtain interval estimates of a_w.

- Propose methodology for soft sensor maintenance and evaluation regarding the need for retraining based on statistical process control.
- Apply the methodology in an application example involving a pet food industry.

1.4 DISSERTATION LAYOUT

The content present in each of the following Chapters of this text are briefly described below:

- Chapter 2 presents the theoretical background and literature review approaching the followed subjects: Water in food, sorption isotherms, soft sensors and machine learning.
- Chapter 3 presents a methodology proposed to build a soft sensor to predict water activity
- Chapter 4 contains the application example in the context of a pet food industry. The proposed methodology is used to build soft sensors using ML techniques to predict a_w from moisture content, which is measured in-line by means of a humidity sensor.
- Chapter 5 presents the conclusion, limitations and suggestion for future works.

2 THEORETICAL BACKGROUND AND LITERATURE REVIEW

In this chapter, the definitions and explanation about the key topics and concepts of this dissertation are presented. Furthermore, a review of the current researches in the area is presented.

2.1 WATER IN FOOD

Water is an important constituent of food that affects its safety, stability, quality and physical properties. There is a large range of water percentage in foods some of them presents a small percentage and others, as liquid foods, it can reache even more than 98% (LEWICKI, 2004). The water usually is evaluated by two basic types of water analysis: water content and water activity.

The concept of water content is the most familiar. It is related to the total amount of water present in a sample. Its determination can be performed using either direct or indirect methods. The direct method may be based on physical separation techniques like drying. It is also called "moisture", once the measure comprises the mass loss due to water and to all the volatile compounds in the sample or formed during its heating (MATHLOUTHI, 2001).

Indirect determination relies on the spectroscopic properties of water molecules. It can be made by Nuclear Magnetic Resonance (NMR), infrared and Raman spectroscopy which are non-destructive techniques, as well as microwave spectroscopy. Interaction of water molecule with electromagnetic radiation may be used in the analysis of water content. The NIR can be used, once absorption of water occurs at different wavelengths. The ratios of the intensities of the bands at 1950 and 1450 nm are used as a measure of water content. Computerised NIR spectrometers are used in different food industries for the determination of water content and food constituents (MATHLOUTHI, 2001).

Water content is also essential to meet product nutritional labeling regulations, to specify recipes, and to monitor processes. However, it is not a reliable for predicting microbial responses and chemical reactions in materials when considered as the unique indicator (METER GROUP, 2018; MOTA *et al.*, 2008).

The physical, chemical and microbiological stability of food depends on the water content and its interaction with food components. Scott (1953) and contributors were the first to establish that is the a_w, and not the water content, that is correlated with microbial growth,

once a_w is the amount of water available for interactions and cross-reactions with other molecules and solutes. When water interacts with solutes, it is not available for other interactions, and hence, the water required for growth of microbes is not available and they get suppressed (PRABHAKAR & MALLIKA, 2014; SABLANI *et al.*, 2007).

The concept of water activity has been used as a reliable assessment of microbial growth, lipid oxidation, chemical reactivity, enzymatic and non-enzymatic activities, which also interferes in other food characteristics as texture and taste. By controlling water activity in addition to factors as pH and temperature, the food storage stability and microbial safety can be controlled (BAŞER & YALÇIN, 2017; DECAGON DEVICES, 2012; LEWICKI, 2004; ROOS, 2003).

The free water is the water available for microbial growth. It is a function of the water vapor pressure in food, which has led to the development of water activity concept. What can be determined by the mathematical model developed by Scott (1953) as the ratio between water vapor pressure into the food (p_w) and pure water vapor pressure (p_{w0}) under the same condition of temperature and pressure (Equation 1). It also can be related with the equilibrium relative humidity (ERH) (FRANKS, 1991; LEWICKI, 2004; NIELSEN *et al.*, 2015; FDA, 2015):

$$a_w = \frac{p_w}{p_{w0}} = \frac{ERH}{100} \quad (1)$$

When a solution or food is placed in an environment, it will lose or gain water until equilibrium is reached and the water activity in the atmosphere and in the food have become equal. This has a great importance in a_w measurements, because a_w is often derived from the equilibrium relative humidity of the headspace of a food system (ROOS, 2003).

The vapor pressure of ideal solutions is related to the mole fraction of water, as defined by Raoult's law (Equation 2), which is based on the observation that the partial vapor pressure of each component in a solution is a function of its mole fraction and the vapor pressure of a pure liquid at the same temperature. In the case of water, assuming that the solution behaves as an ideal solution, Raoult's law is linear. However, often the vapor pressure deviates from the ideal behavior as solute concentration increases. For a real solution, an activity coefficient (γ_w) is defined, and the water activity is related to the mole fraction of water, x_w, by Equation 3. Therefore, water activity may also be considered as the

'effective' mole fraction of water. However, food systems are rather complex, the water activity is difficult to estimate, and must be measured for most foods (ROOS, 2003).

$$p = x_w p_{w0} \quad (2)$$

$$a_w = \frac{p_w}{p_{w0}} = \gamma_w x_w \quad (3)$$

2.1.1 Water activity and food preservation

The range of a_w value is from 0 to 1, where the greater the a_w is the greater quantity of free water available to microorganism make its reactions for self-development. Depending of microorganism it is required a range of a_w for its development. Molds require for optimum growth an a_w within 0.60-0.88. Many microorganisms prefer a_w values of 0.99. Most spoilage bacteria need a_w higher than 0.91, whereas spoilage molds can grow even at a_w = 0.80. Many types of yeasts can grow at level of 0.80 in foods like fermented sausage, dry cheese or in products with 65% sucrose or 15% salt (BAŞER & YALÇIN, 2017; PRABHAKAR & MALLIKA, 2014).

When crispness and crunchiness are required in food products a_w values below 0.65 are usually maintained. These values contribute to limit moisture migration in composite food products and can also be used to predict the migration of moisture that affects food characteristics. Water activity can be decreased as temperature and pressure increasing. Manipulating properly a_w levels in food products, their shelf life can also be prolonged. The addition of certain agents or solutes like glycerol, sucrose, sodium chloride, among others, can lower a_w and increase the storage period. These agents bind most of the available water molecules, thus microorganisms cannot access them for growth purpose. The properly manipulation of a_w also contribute with demanding requirements for packaging and the preservation of stored foods, offering greater flexibility in food-processing and distribution-marketing operations (PRABHAKAR & MALLIKA, 2014).

Within crispy or crunchy foods, which present low water activities, the interactions between water and macromolecules are so strong that water is devoid of solvent properties. Under these conditions, mobility of small molecules as well as macromolecules is significantly reduced. The increase of water content can lead to either a plasticizing or an anti-plasticizing effect. Water acquires solvent properties and enables the mobility of polymer chains. This plasticization of polymer chains facilitates deformation. Material becomes soft,

extensible and flowable and losses crispness, hardness or toughness. Thus, it may not meet specifications or may not be accepted by consumers (LEWICKI, 2004).

2.1.2 Methodologies to measure a_w

The measurement of a_w is based on the colligative properties of solutions and may be measured by a number of means, two of them are (PRIOR, 1979; TROLLER, 1983):

1. Measurement of the freezing point depression of a solution and conversion to a_w:

The lower freezing temperature phenomenon is called the freezing point depression in the case of water it happens at 0 °C. The freezing point is lowered due to the presence of solutes, solids, in the water. Which is primarily dependent on the concentration or amount of solids. In this technique, the depression in freezing point determines the effective gram moles of solutes that are present (Equation 4 and 5). This value is then used, by means of one form of Raoult's law, to arrive at the a_w (Equation 3), assuming the activity coefficient equal to one. Observing that it is not applicable for solid foods.

$$\mathrm{n_2} = \frac{\text{grams solvent}}{1000} . \frac{\Delta \mathrm{T_f}}{1.86} \quad (4)$$

$$\Delta \mathrm{T_f} = \mathrm{T_f(solvent)} - \mathrm{T_f(solution)} \quad (5)$$

Where, n_1 is moles of water in system, n_2 is moles of theoretical solute in system, ΔT_f measures the freezing point depression in °C, and T_f is the freezing point.

2. Measurement of the ERH of a solid or liquid. This may be determined by two means:

(2.a) a quantity of the substance to be measured is enclosed with a small quantity of gas (usually air) and the relative humidity or vapor pressure of this gas is measured once equilibrium is reached. Where the relative humidity is the amount of water present in the air compared to the greatest amount it would be possible for the air to hold at certain temperature.

(2.b) a sample of the substance is placed in a gas at known temperature and relative humidity and the moisture absorbed or lost by the sample is determined.

Two different types of a_w instruments are commercially available. One uses chilled mirror dew point technology and the other measures relative humidity with sensors that change electrical resistance or capacitance. Each has advantages and disadvantages. The methods vary in accuracy, repeatability, speed of measurement, stability in calibration, linearity, and convenience of use (PRABHAKAR & MALLIKA, 2014).

The measurement of relative humidity from the dew point has been used for decades. In such a technique, the sample is placed in a sealed chamber containing a mirror and sensors to detect condensation. When equilibrated, the relative humidity of the air in the chamber is the same as the water activity of the sample. The major advantages of the chilled-mirror dew point method are accuracy, speed, ease of use, and precision. The range of a_w may be from 0.030 to 1.000, with a resolution of 0.001 and accuracy of 0.003. Measurement time is typically less than 5 minutes (FERBLANTIER *et al.*, 2003; PRABHAKAR & MALLIKA, 2014).

Other a_w sensors use a resistance or capacitance to measure relative humidity. The sensor is composed of a hygroscopic polymer and has an associated circuit to provide the value of ERH. They are less expensive, but they are not as faster or as accurate as the sensors based on the chilled-mirror dew-point method. Since these instruments relate an electrical signal to relative humidity, the sensor must be calibrated with known salt standards. In addition, the ERH is equal to the sample a_w only if the sample and sensor temperatures are the same. Some capacitive sensors need between 30 and 90 minutes to come to temperature and vapor equilibrium. Accurate measurements with this type of system require good temperature control. These instruments measure over the entire a_w range (0 to 1), with a resolution of 0.005 and accuracy of 0.015 (FERBLANTIER *et al.*, 2003; PRABHAKAR & MALLIKA, 2014).

2.1.3 Equipment to measure a_w

Among the equipment already commercialized to make the measurement from small samples, there are (METER FOOD, 2018, ALFAMARE, 2018; SOURCE TECHNOLOGY, 2018):

- Aqualab Series 4TE and 4TEV: they are water activity meters per dew point and capacitance, capable of measuring all types of samples (solid, semi-solid and liquid).
- Aqualab 4TE Duo: it is able to measure a_w and moisture content, the second one is determined using the measured a_w in a pre-loaded product-specific isotherm model.
- Aqualab PRE: it uses a sensor with dielectric constant and a system that measures the vapor pressure in a hermetic chamber.
- Aqualab TDL: it measures the relative humidity of the air with an infrared beam, one of its advantages is that the reading is not compromised by volatile compounds.

- Pawkit: it is a portable and compact meter, which uses a capacitive sensor to make a_w determination in less than 5 min. The sensor converts the humidity value into a specific capacitance, which is then measured electronically by the circuit. This signal is then translated by the software and displayed as water activity on the instrument screen.
- Automated Quality Analysis - AQA^{TM}: it is a central unit, placed in a location where a product can be fed from multiple conveyor belts if necessary. It consist of a movable sample cup, which transports the product to the different analysis station to measure moisture content, water activity, and density.

The Aqualab equipment presented above were projected to use in laboratory, the Pawkit is portable but it can't be used in any environment, as the others equipment it's necessary a place without great variation of temperature. The read time of these equipment is approximately of 5 min. The AQA is the only one that was project to collect the sample and make the analysis automatically, the water activity is measured within 3-6 min depending of the product (METER, 2018; SOURCE TECHNOLOGY, 2018). The AQA demands some investment and changing in the production line when it wasn't installed before the line be already in operation.

2.1.4 Previous works involving a_w measurement

After a review in the literature it was possible to see works, for a variety of food, proposing new methodologies for measuring a_w, through its relationship with other parameters, aiming its online monitoring. For instance, Traffano-Schiffo *et al.* (2015) made a study of application of dielectric spectroscopy to predict the a_w of meat during drying process, appearing to be a promising method for real-time monitoring. This work showed that there is a direct relationship between the dielectric loss factor at 20 GHz, considering the number of water molecules on the surface of the sample. What allowed to determine the point of water activity at which the surface temperature reaches the air-drying temperature, by using the loss angle at 20 GHz. At this point of water activity, the liquid phase disappears, and internal transport controls the drying process. As a result, it was developed a useful tool to predict the surface water activity by using the loss angle at 20 GHz for meat.

Also related with the meat industry, in 2007, Muñoz *et al*. developed a fuzzy controller for sausages' drying from a_w monitoring. The a_w was estimated on a sausage surface using its surface temperature, the air's temperature and relative humidity. The authors showed that the controller was able to adapt to the changing conditions of the dryer, which makes it suitable for drying monitoring when the factors affecting the process cannot be controlled directly. Collell *et al*. (2010), reported a technique using Fourier Transform - NIR (Near Infrared) spectroscopy to predict moisture and NaCl contents on surface of fermented pork sausages, they also suggest that it could also be useful to predict a_w. This technique proposed the construction of various PLS regression models correlating NIR pre-processed spectral data with chemical results of moisture and NaCl contents and a_w. The results showed as appropriate for online applications to monitor non-destructive drying process.

A technology related to NIR, called Near Infrared Hyperspectral Chemical Imaging was presented, by Achata *et al.* (2015), as potential to predict moisture content and a_w in low moisture systems (coffee, wafer, soybean). The chemical image from the near infrared spectrum is a technique to spatially characterize the spectral properties of samples. Due to the rapid acquisition of chemical images, many samples can be evaluated simultaneously, providing the potential for online evaluation of samples during processing. The obtained results indicate the suitability of NIR spectral image combined with chemometrics for such a task and also to provide a visualization of the spatial distribution of the moisture content and a_w in the analyzed food samples.

In the case of pet food industry, Pedersen and Rasmussen (2015) developed a prototype of a new system to make automatic in-line sampling and measurement of moisture, density, and a_w The work's aim was to improve the measurement speed and minimize problems related to the handling of samples. It was aimed a_w measurement range from 0.2 to 0.8, with uncertainty between 0.02 and 0.05, and the time of measuring less than five minutes. This work was tested with pet food and showed satisfactory results, however, in this case still there is a delay to provide a_w values.

Concerning works that use ML techniques, Zhang and Mittal (2013) learn the isothermal data and determine the relationship between moisture content in equilibrium and water activity for various temperatures and food products, using Artificial Neural Networks. Data from 53 food products were taken from the literature, input information used was temperature, relative humidity. The mean relative error of prediction was about 2% and the

optimum structure identified was 3 intermediate layers, each with 30 nodes. The authors also compared the neural network model with eight existing mathematical models and concluded that the neural network model can be used with comparable accuracy to predict the equilibrium moisture content.

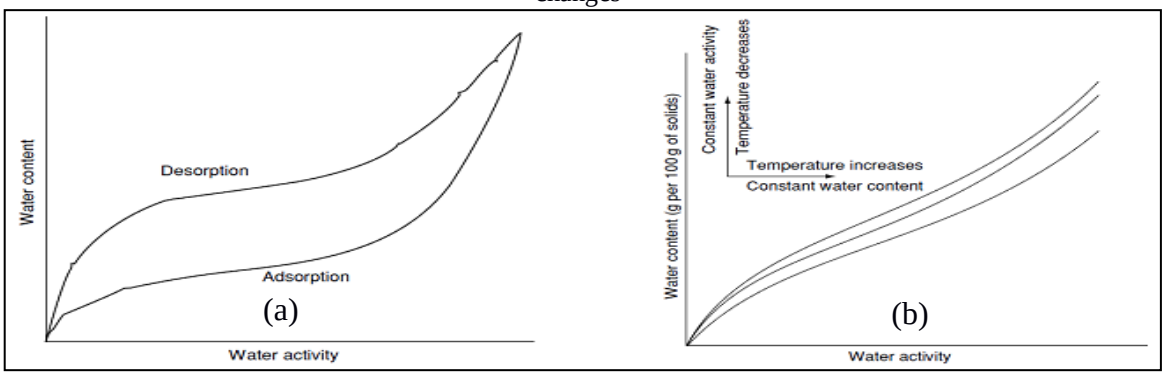
Neural networks were also used in conjunction with fuzzy inference system (ANFIS) to model sorption isotherms and predict moisture content at equilibrium for milk powder, as reported by Simha *et al.*(2016). According to the authors, the used method has shown to be a simple and efficient technique to predict complex non-linear food relations.

2.2 SORPTION ISOTHERM

The state of water in solids can be graphically or analytically represented by the so-called sorption isotherms. The sorption isotherm shows a nonlinear relationship between water content and water activity within the material in the state of equilibrium at a constant temperature. It describes in a specific water content, the balanced ratio between an amount of water sorbed by components of the biological material and the vapor pressure or relative humidity at a given temperature. This ratio depends on the chemical composition of the food such as fat, starch, sugar, protein, among others (FDA, 2015; LEWICKI, 2004; PARK *et al.*, 2008; PRABHAKAR & MALLIKA, 2014).

Figure 1 (a) presents the general shape of a sorption isotherm. It is possible to see that an isotherm can present two shapes: adsorption isotherms, obtained when a totally dry food is placed in atmospheres with several increments of relative humidity and the weight increase of the food is due to the gain of water during the hydration process; desorption isotherm, obtained when the initially moist food is placed under the same atmospheric conditions created for adsorption and in this case the decrease in weight of the food is caused by the loss of water during the dehydration. These adsorption and desorption curves do not coincide, and this difference between the isotherms is known as the hysteresis phenomenon In general, the a_w of a food with a constant water content increases with increasing temperature, while the water sorption decreases with increasing temperature at the same a_w, as illustrated in Figure 1-(a) General representation of sorption isotherm curve (b) General behavior when the temperature changes(b). However, the temperature dependence may vary with the moisture and water contents present in food for foods with a_w above 0.65 (PARK, 2008; ROOS, 2003).

Figure 1- (a) General representation of sorption isotherm curve (b) General behavior when the temperature changes



Source: Roos (2003)

Two basic methods can be used to obtain the constant temperature sorption curves. In the first method, the food of known moisture content is allowed to reach equilibrium with a small empty space in a limited enclosure, and the partial pressure of a_w is measured manometrically, or the relative humidity is measured using a hygrometer (FDA, 2015).

A second basic method for the preparation of isotherms is the exposure of a small sample of food to various atmospheres/environments of constant humidity. Once equilibrium is reached, the moisture content is determined gravimetrically or by other methods. Several saturated salt solutions are available for this purpose. Saturated salt solutions have the advantage of maintaining a constant moisture as long as the amount of salt content is above the saturation level.

There are mathematical models for modeling the isotherms as the Brunauer, Emmett and Teller (BET) sorption isotherm, which was the model that had the greatest application to determine water sorption of foods, in the past. However, it was valid only for a limited range of a_w, up to only 0.3 - 0.4. The BET monolayer concept was found to be a reasonable guide with respect to various aspects of interest in dried foods, in spite of the theoretical limitations of the BET adsorption analysis. More recently, the Guggenheim, Anderson and de Boer (GAB) isotherm equation has been widely used to describe the sorption behavior of foods. Having a reasonable small number of parameters (three), this is considered to be one of the most used models to better represent experimental data within a range of a_w of greater practical interest in food, that is, between 0.1 and 0.9 (TIMMERMANN *et al.*, 2001).

The knowledge of the sorption behavior of foods is useful in the processes of concentration and dehydration because (FDA, 2015):

- It is important in the design of the processes themselves; because it has a major impact on the facility or difficulty of water removal, which depends on the partial pressure of water on food and the energy of binding water in food.
- Water activity affects food stability and therefore must be brought to an adequate level at completion of drying and maintained within an acceptable range of values during storage.

2.3 SOFT SENSORS

The rapid advance in sensors, computers, communication and information technologies provide significant changes in the configuration of factories. This includes fully automated quality inspection, which can process large amounts of measurement data. Other production-related activities and business functions will also be integrated into the company's information management network, which ensures real-time access to critical production data for better decision making. This new approach is called electronic quality control (e-quality control), and one of its benefits is the ability to predict variations, performance losses during the various stages of production. This means that the traditional quality control scheme, which depends on sampling techniques, would be replaced by automated, sensor-based inspection methods that provide the unprecedented level of data processing and handling. In modern production environments, large amounts of data are collected through the database management systems and data warehouses of all the areas involved, such as assembly, product and process design, material control and planning, entry and scheduling of orders, recycling, among others (TSENG *et al.*, 2016).

In 1990 and early 2000, researches started to use a large quantity of data being measured and stored in the process industry by building predictive models based on this data, they were called soft sensors (SS), what is a combination of the words “software”, because the models are usually computer programs, and “sensors”, once the models are delivering similar information as their hardware counterparts. SS are defined as inferential models that use online available sensor measurements for online estimation of quality variables, which cannot be automatically measured, or can only be measured at high cost, sporadically, or with high delays (e.g. laboratory analysis) (KADLEC *et al.*, 2009; SOUZA *et al.*, 2016).

In order to measure the quality variables in real-time, computational intelligence methodologies can be used to build intelligent/computational sensors to infer the value or the

quality target variables from other measured process variables. The basis for building such intelligent sensors is that the values of the target variable, or the product quality, have a functional relationship with other process variables that can be measured online. Such kind of intelligent sensor is one of the applications of soft sensors (KADLEC *et al.*, 2009; SOUZA *et al.*, 2016).

In general, two main classes of SS are identified: model-driven and data-driven. The first one, also called white-box model, is based on First Principle Models (FPM), what means that they have full knowledge about physical and chemical background of the process. These models usually have focus on the description of ideal steady-states of the processes. The second one, also called black-box model, is based on the data measured within the processing plants, the model itself has no knowledge about the process, it is based on empirical observation and thus describe the true conditions of the process in a better way and can be used in more complicated industrial processes. Between these two extremes, there are models called gray-boxes, which are a combination of the model-driven and data-driven SS (GE, 2017; KADLEC *et al.*, 2009).

2.3.1 Soft Sensor Regression (SSR)

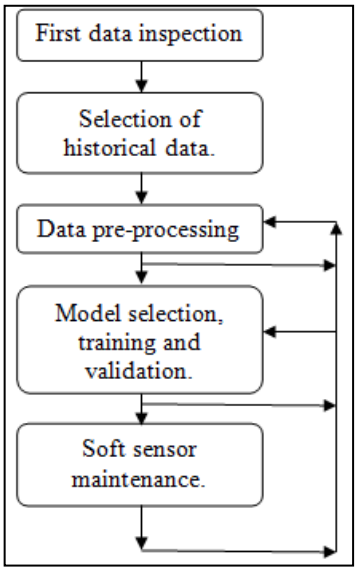
An SSR is a regression model which uses easy-to-measure variables to predict a hard-to-measure variable. It is the subject of research in many areas, as chemometrics, which use statistical methods for extracting information from data sets that often consist of many measured variables. This task demands knowledge of statistics, numerical analysis, operation analysis, and applied mathematics. In chemometrics literature, it's possible observe the use of different SSR approaches including machine learning and pattern recognition, artificial intelligence, system identification, and statistical learning theory. Even though the objectives and emphasis on all these areas are different, they are intrinsically connected by the necessity to learn models from data (SOUZA *et al.*, 2016).

SSR development encompasses the same design cycle of classical regression systems. However, it has its own peculiarities. Figure 2 summarises the stages to build a soft sensor.

In the first stage, the data are collected, and its goals include handling of missing data and outliers. The objectives of the second stage are the selection of the most relevant inputs, and possibly the respective time lags. The model choice and training require the correct selection and learning of the model. Building a linear model should be considered before a

nonlinear, if the first does not show satisfactory results, then nonlinear model should be taken into account. In this step a lot of machine learning techniques have been applied. The model validation step is necessary to judge if the learned model reproduces the target variables within acceptable quality or performance levels. The last step is the SS maintenance, where the goal is to maintain a good SSR response in the presence of process variations or some data change. This step is necessary due to the fact of changes of the data may deteriorate the SS's performance, being necessary to compensate by adapting or re-developing its model (KADLEC *et al.*, 2009; SOUZA *et al.*, 2016).

Figure 2-Methodology for soft sensor development



Source: Kadlec *et al.*(2009)

The main application of SS is for predicting process variables that can be determined either at low sampling rates or through off-line analysis. Some works that used it with machine leaning applied to food industry:

- CHAURASIA *et al.* (2018) brought a recent study, in the food industry, about the performance of different models in predict drying time of mosambi peel in function of temperature, salt concentration and thickness of drying bed. The applicability of Gaussian process regression (GPR)-based approach for modeling drying kinetics was analyzed, compared with the commonly used approaches, one of them was artificial neural network (ANN). The models were

validated by comparing model simulations with observed values for unseen data. The models were compared based on performance indices like coefficient of determination, mean square error, root mean square error (RMSE), model predictive error, mean average deviation, goodness of fit, and chi-square analysis. The study concluded that regression-based models of the Gaussian process are efficient in modeling drying kinetics and have better prediction capability than ANN.

- Artificial Neural Networks were also used to develop a soft sensor for on-line estimation of ethanol concentrations in wine stills, as showed in the paper wrote by Osorio *et al.* (2008). The soft sensor was calibrated with data from laboratory and industry, and the prediction is based in four temperature measurements in the still. An important additional advantage of the soft sensor pointed by the authors was that it could reduce delays in ethanol measurements, which should in turn improve automatic control loop performance.
- Bruwer *et al.* (2007), presented a soft sensor to predict textural properties from a commercial snack using vibrational signals acquired “passively” from the process using accelerometers and/or microphones. The data were used to predict two pseudo-texture properties, "blister level" and "brittleness", that could predict approximately 90% of the variation in the larger set of measurements. The prediction of these two properties would enable real-time multivariate statistical process control, as well as the possibility of automated feedback control of the product quality.

2.4 MACHINE LEARNING

Nowadays the large amount of data collected far exceed our capacity to digest and interpret them. Machine Learning (ML) can help with these applications by providing tools to clean up, filter, and identify the most important data subset. These algorithms learn information directly from the data without having a predefined model or equation relating inputs to outputs, using for that the statistical theory to build mathematical models. These model may be predictive to estimate values of the output variables based on unseen input values, or descriptive to gain knowledge from data, or both (ALPAYDIN, 2010; ESCOBAR & MORALES-MENENDEZ, 2018; PORTER *et al*., 2013).

ML algorithms can be classified in six categories, based on underlying mappings between input data and output data during the learning phase: supervised learning, unsupervised learning, semi-supervised, transductive inference, inductive inference and reinforcement learning. The two first categories are better known (AWAD & KHANNA, 2015).

Supervised learning algorithms are techniques that extract associations between independent attributes and a designated dependent attribute. In this case a set of examples are presented in the form: input (x_i) and output (y_i), denoted by label. Given a set of data labeled in the form (x_i, y_i) one must produce a model, capable of accurately predicting the label of new entries that were not presented previously. The process of inducing a model from a data sample is called training, and a model is essentially a function *f(x)* that receives the value *x*, and returns the value *y*. A function model based on a supervised learning algorithm, when well-trained, can predict with accuracy the labels for unfamiliar or unseen data instances. The goal of learning algorithms is to minimize the error for a given set of inputs (the training set). The errors – deviation between predicted and actual values – are summarized by a cost function, which measures the overall accuracy of a model. The performance of models developed using supervised learning depends upon the size and variance of the training data set to achieve better generalization and greater predictive power for new data sets. The labels represent the phenomenon of interest on which to make predictions, when the labels assume discrete values, there is a classification problem. If the labels have continuous values, there is a regression problem (AWAD & KHANNA, 2015; LORENA & CARVALHO, 2007).

In the case of unsupervised learning, no examples of labels are provided. The algorithm learns to group the entries according to a distance metric. The main objective is to find patterns or tendencies that help in the understanding of the data (LORENA & CARVALHO, 2007).

There are many machine learning algorithms (Decision tree, Random Forest, Support Vector Machine, Neural Network, so on), most of them are complex. For each ML algorithm there are two types of model parameters: ordinary parameters that are automatically optimized or learned during model's training and hyper-parameters that are typically set manually by the user before an ML model is trained. The model building process is complete when the model achieves satisfactory prediction accuracy. Otherwise, changes on the hyper-parameters values are necessary and also model retraining. This process is repeated until a model with

satisfactory accuracy is obtained or it cannot be improved much further more. There is an enormous number of possible combinations of algorithms and hyper-parameter values, the model building process can easily take thousands of manual iterations (LUO, 2016).

However, in the literature is available several hyper-parameter selection approaches, grid search and cross validation are the most used. The grid search refers to the exhaustive search process over a subset of the workspace, and its goal is to find points where accuracy is as large as possible. However, it has as a disadvantage the processing time, as well as the delimitation of the space to be investigated when the data is not previously known. Cross-validation is a technique used to evaluate the generalization capacity of an algorithm when exposed to a new data set. It defines a dataset that is used for testing the trained model during the training phase for over-fitting. It can also be used to evaluate the performance of various prediction functions. In k-fold cross-validation, the training dataset is arbitrarily partitioned into *k* mutually exclusive subsamples (or folds) of equal sizes. The model is trained *k* times, where each iteration uses *k-1* subsamples for training, and the remaining subsample is applied for testing the model (cross validation). The *k* cross-validation results are calculated to estimate the accuracy as a single estimate. Another approach is the use of heuristics and meta-heuristics such as genetic algorithms, differential evolution, particle swarm optimization, among others (ACOSTA *et al.*, 2016; AWAD & KHANNA, 2015).

2.4.1 Support Vector Machines (SVM)

SVM, also known as Kernel Machines, belongs to the supervised learning techniques. Whose learning problem setting is as follows: there is some unknown and nonlinear dependency (mapping, function) $y = f(x)$ between some high-dimensional input vector x and scalar output y (or the vector output y as in the case of multiclass SVMs). There is no information about the underlying joint probability functions. Thus, one must perform a distribution-free learning. The only information available is a training data set $\{(x_i, y_i) \in X \times Y\}$, where i is the number of the training data pairs and is therefore equal to the size of the training data set. Although this problem is very similar to the classical statistical inference, the SVM can be appropriate to solve many contemporary problems, considering the facts that modern problems are high-dimensional, and the underlying real-life data generation laws may typically be very far from the normal distribution. Unlike what occurs in the in classical statistical inference the maximum likelihood estimator (and consequently the sum-of-error

squares cost function) should be replaced by a new induction paradigm that is uniformly better, in order to model non-Gaussian distributions (WANG, 2005).

SVMs are the so-called nonparametric models, their "learning" (selection, identification, estimation, training or tuning) is the crucial issue. Unlike in classic statistical inference, the parameters are not predefined, and their number depends on the training data used. In other words, parameters that define the capacity of the model are data-driven in such a way as to match the model capacity to data complexity (WANG, 2005).

Besides the fact that the algorithm can be use to solve classification and regression problems. Another highlight is its capacity in handling nonlinear problems, what is possible through the use of Kernel functions whose main objective is to map input vectors to a high-dimensional feature space, in which the decision or regression function is linear (BURGES, 1998; SMOLA & SCHÖLKOPF, 2004; VAPNIK & CORTES, 1995). The different Kernel functions will be seen further ahead.

Due to these advantages, this technique has a wide range of applications, being in works related to regression problems as determination biodiesel content in diesel fuel, estimation of biomass concentration, prediction of quantity of dissolved oxygen, biomass, substrate and penicillin concentrations in batch bioprocess, quantification of adulterants in milk powder (CESAR *et al.*, 2013; FERRÃO *et al.*, 2007; LIU *et al.*, 2010; YU, 2012). Being also applied in classification problems related to cancer diagnosis, bioinformatics, survival analysis, food quality control, metabolism analysis, tea classification, among others (LI *et al.* 2009a; LUTS *et al.*, 2010; SMOLA & SCHÖLKOPF, 2004).

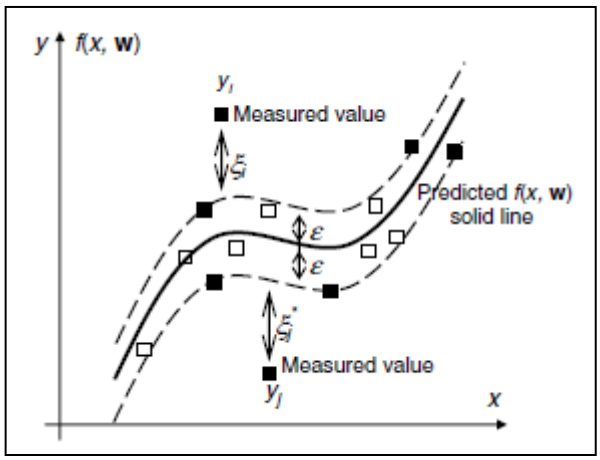
2.4.1.1 Support Vector Regression (SVR)

The basic idea of the SVR is that if we have a set of training data $\left((x_1, y_1);\ (x_2, y_2); \ldots; (x_n, y_n)\right) \in R^n$ x R , the goal is to find the function *f(x)* that presents at most an ε deviation of the real targets y_i for all training data. That is, errors are not taken into account if they are smaller than ε, but any deviation greater than ε is not accepted. For the case of a one-dimensional example, displayed in the Figure 3, the continues-valued approximated can be written as Equation (6):

$$y = f(x) = \langle w, x \rangle + b = \sum_{i=1}^{n} w_i . x_i + b \qquad (6)$$

Where $y, b \in \mathbb{R}$ and $x, w \in \mathbb{R}^n$; w, x and b are the vector of weights, the input vector and the bias, respectively.

Figure 3-One-dimensional linear SVR



Source: Wang (2005)

This function approximation problem is formulated as an optimization problem that attempts to find the narrowest tube centered around the surface, while minimizing the distance between the predicted and the desired outputs. Thus, the SVR training problem is given by Equation 7, where $\|w\|$ is the magnitude of the normal vector to the surface that is being approximated (ACOSTA *et al.*, 2016; AWAD & KHANNA, 2015).

$$\text{Minimize } \frac{1}{2}\|w\|^2 \tag{7}$$

The magnitude of w acts as a regularizing term and provides optimization problem control over the flatness of the solution. SVR adopts an ε-insensitive loss function, penalizing predictions that are farther than ε (tolerance) from the desired output. As it is possible to see in the figure above the value of ε determines the width of the tube; a smaller value indicates a lower tolerance for error and also affects the number of support vectors (points of the sample set that have relevant role used in the estimated regression function) and the solution sparsity. If ε is decreased, the boundary of the tube is shifted inward. Therefore, more data-points are around the boundary, which indicates more support vectors. Similarly, increasing ε will result in fewer points around the boundary. Because it is less sensitive to noisy inputs, the ε-insensitive region makes the model more robust. The loss functions should be convex to ensure that the optimization problem has a unique solution that can be found in a finite

number of steps. The Equation 8 presents a linear loss function (AWAD & KHANNA, 2015; SMOLA & SCHÖLKOPF, 2004):

$$L_{\varepsilon}\big(y, f(x,w)\big) = \begin{cases} 0, & |y - f(x,w)| \leq \varepsilon; \\ |y - f(x,w)| - \varepsilon, & \text{otherwise} \end{cases} \quad (8)$$

Sometimes this may not be the case and it is also possible to exist unpredictable restrictions of the optimization problem. Therefore, to solve this issue, slack variables (ξ, ξ^*) are introduced so that noises, outliers and points outside the region between - ε and + ε can be present in the available training data set.

Thus, the optimization problem in Equation 9 is obtained; C is a regularization, a tuneable parameter that gives more weight to minimizing the flatness, or the error. For example, a larger C gives more weight to minimizing the error.

$$\text{Minimize } \frac{1}{2}\left|\left|w\right|\right|^2 + C\sum_{i=1}^{n} \xi_i + \xi_i^* \quad (9)$$

$$\text{Subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

This constrained quadratic optimization problem can be solved by finding the Lagrangian (see Equation 10).

$$\begin{aligned} L = \frac{1}{2}\left|\left|w\right|\right|^2 + C\sum_{i=1}^{n} \xi_i + \xi_i^* - \sum_{i=1}^{n} (\eta_i \xi_i + \eta_i^* \xi_i^*) \\ - \sum_{i=1}^{n} \alpha_i (\varepsilon + \xi_i - y_i + \langle w_i, x_i \rangle + b) \\ - \sum_{i=1}^{n} \alpha_i{}^* (\varepsilon + \xi_i{}^* - y_i + \langle w_i, x_i \rangle + b) \end{aligned} \quad (10)$$

Where L is the Lagrangian, $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ are Lagrange multipliers, and need to be positive. The minimum of Equation 10 is found by taking its partial derivatives with respect to the variables and setting them equal to zero, based on the Karush-Kuhn-Tucker (KKT) conditions. The partial derivatives with respect to the Lagrange multipliers return the constraints, which have to be less than or equal to zero, as illustrated in Equations 11, 12 and 13.

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) = 0 \quad (11)$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) x_i = 0 \quad (12)$$

$$\frac{\partial L}{\partial \xi_i^{(*)}} = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \quad (13)$$

Substituting (11), (12) and (13) into (10) yields the dual optimization problem.

$$\text{Maximize } -\frac{1}{2}\sum_{i,j=1}^{n} (\alpha_i^* - \alpha_i) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i (\alpha_i + \alpha_i^*) \quad (14)$$

Subject to $\sum_{i,j=1}^{n} (\alpha_i^* - \alpha_i)$ and $\alpha_i^*, \alpha_i \in [0, C]$

The equation 13 is reformulate as $\eta_i^{(*)} = C - \alpha_i^{(*)}$, and the equation 12 is re-written as $w = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) x_i$, changing the Equation 6, the followed equation 15 is obtained:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \quad (15)$$

It's possible to compute the parameter *b* by exploring the KKT conditions, which state that at the optimal solution the product between dual variables constrains has to be equal to zero. What brings to

$$\begin{aligned} &\alpha_i (\varepsilon + \xi_i - y_i + \langle w_i, x_i \rangle + b) = 0 \\ &\alpha_i{}^* (\varepsilon + \xi_i{}^* - y_i + \langle w_i, x_i \rangle + b) = 0 \end{aligned} \quad (16)$$

$$\begin{aligned} &\eta_i \xi_i = (C - \alpha_i) \xi_i = 0 \\ &\eta_i^* \xi_i^* = (C - \alpha_i{}^*) \xi_i{}^* = 0 \end{aligned} \quad (17)$$

Some conclusions can be taken: only samples (x_i, y_i) with corresponding $\alpha_i{}^{(*)} = C$ lie outside the ε-insensitive tube. Secondly, there can never be a set of dual variables which are both simultaneously nonzero. Third, for support vectors corresponding to $\alpha_i{}^{(*)} \in (0, C)$, $\xi_i{}^{(*)} = 0$.

Then, with some manipulation *b* can be computed as presented in Equation 18:

$$\begin{aligned} &b = y_i - \varepsilon - \langle w_i, x_i \rangle \text{ for } \alpha_i \in [0, C] \\ &b = y_i + \varepsilon - \langle w_i, x_i \rangle \text{ for } \alpha_i{}^* \in [0, C] \end{aligned} \quad (18)$$

When a nonlinear model is required to adequately model the data, it is necessary to extend the linear ε-SVR to non-linear regression (ACOSTA *et al.*, 2016; LI *et al.*,2009). All

the explanation and steps presented before can be applied to a non-linear function, and the Equation 6 is presented as Equation 19:

$$f(x) = \langle w, \varphi(x) \rangle + \text{b} \quad (19)$$

Where $\varphi(.): \mathbb{R}^n \rightarrow \mathbb{R}^{n_i}$ is the nonlinear function that maps the input space to a high dimensional feature space where the linear regression is performed, and the dimension n_i is this space can be infinite. The use of a kernel function enables the original input data to be mapped into the high-dimensional feature space, where a linear regression can be used

2.4.1.2 Kernel function

The kernel function has an paramount importance. It serves to increase the number of dimensions and transform linearly inseparable data into linearly separable data. In addition, the function is also a mathematical trick to calculate the dot product in the input space, which is a significant step during the training of SVM model (LI *et al.*, 2009).

A kernel, K, is a function that takes two points (x_i and x_j) in the original input space and implicitly calculates their dot product. In this way, the kernel can transform a set of non-linearly separable data into a set of linearly separable data (LI *et al.*, 2009). The most commonly used kernel functions are linear, polynomial, Gaussian radial basis function (RBF) and sigmoid, which are presented in the Table 1 with their respective parameters and constants:

Table 1- Kernel function types

Kernel Function	**Equation**
Linear	$K(x_i, x_j) = a{x_i}^t x_j + b$
Gaussian radial basis function (RBF)	$K(x_i, x_j) = exp\left(-\gamma \left\| x_i - x_j \right\|^2\right)$
Polynomial	$K(x_i, x_j) = \left(a x_i^t x_j + b\right)^n$
Sigmoid	$K(x_i, x_j) = tanh\left(a{x_i}^t x_j + b\right)$

Source: Li *et al.* (2009)

For the linear and sigmoid kernels, a and b are also constants. In the RBF, the parameter γ controls the flexibility of the kernel function. Small values of γ allow the adjustment of all the labels with the risk of over-fitting. On the other hand, large values of γ reduce the kernel to a constant function, making the learning process difficult (ACOSTA *et al.*, 2016; LI *et al.*, 2009).

2.4.1.3 Least Squared Support Vector Machine (LS-SVM)

The LS-SVM method is very close to that of SVR, but the more usual sum of the squares of the errors is minimized, and no ε-based selection is made between samples. Differently from SVM, the principles of LS-SVM involves equality instead (Equation 21) of inequality constraints (Equation 9). The final model may be more accurate and may with less computational effort, once LS-SVM uses the least squares loss function instead of the e-insensitive loss function. In this way, the solution follows from a linear KKT system instead of a computationally hard quadratic problem. Therefore, optimization is easier and computing time is shorter (WANG & HU, 2005).

Analogously to parameter C in the SVR model, LS-SVM has a parameter γ that controls the smoothness of the fit. Similar to SVM method is necessary to minimize a cost function (C) that has a penalization of regression errors according to Equation 20 subject to Equation 21 (BALABIN & LOMAKINA, 2011; FERRÃO *et al.*, 2007; WANG & HU, 2005):

$$C = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2 \tag{20}$$

Subject to $y_i = w^T \varphi(x_i) + b + e_i$

where $\varphi(x_i)$ is the mapping function and e_i are error variables, $\gamma \geq 0$ and must be defined a priori by the user (FERRÃO *et al.*, 2007; WANG & HU, 2005). The Lagrangian is given by Equation 21:

$$L = \frac{1}{2} \|w\|^2 + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \alpha_i \{w^T \varphi(x_i) + b + e_i - y_i\} \tag{21}$$

With the Lagrangian multipliers belonging to $\mathbb{R}$. The optimal result is obtained making the partial derivate equal to zero and then obtaining:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{N} \alpha_i \varphi(x_i) = 0 \tag{22}$$

$$\frac{\partial L}{\partial e_i} = \gamma\, e_i - \alpha_i = 0 \tag{23}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{N} \alpha_i = 0 \tag{24}$$

$$\frac{\partial L}{\partial \alpha_i} = e_i - \mathrm{y_i} + \langle w, \varphi(x_i) \rangle + b = 0 \tag{25}$$

Then combining eq. 22 and 23 and solving for *w*:

$$w = \sum_{i=1}^{N} \alpha_i \phi(x_i) = \sum_{i=1}^{N} \gamma e_i \, \varphi(x_i) \tag{26}$$

Replacing a positive nuclei of transformation $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$ in the original regression equation 21, it is obtained:

$$y_i = \sum_{i=1}^{N} \alpha_i \varphi(x_i)^T \, \varphi\left(x_j\right) + b = \sum_{i=1}^{N} \alpha_i \langle \varphi(x_i), \varphi\left(x_j\right) \rangle \tag{28}$$

2.4.2 Artificial Neural Network (ANN)

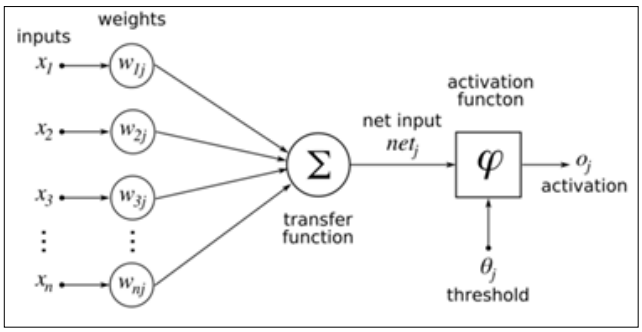
ANN is defined as a predictive tool used to construct mathematical models for a complex system. All patterns presented in experimental data or mathematical descriptions can be learned by ANN. The artificial neural network has been used as an additional tool for various purposes. Activities such as regression and classification are the most common, but it is also used in tasks such as control, modeling and forecasting. As a consequence of the increased popularity of ANN, a number of software tools have been developed to train this type of network, and it is necessary to evaluate the one that best adapts to the problem to be solved (BAPTISTA & MORGADO-DIAS, 2013).

Its best-known class is multi-layer perceptron (MLP). It was primarily shown by Hornik *et al.* (1989) that the multilayer perceptron can be trained to approximate virtually any smooth, measurable function. Unlike other statistical techniques, MLP makes no prior assumptions concerning the data distribution. It can model highly non-linear functions and can be trained to accurately generalize when presented with new, unseen data. These features of MLP make it an attractive alternative to develop numerical models, and also when choosing between statistical approaches (GARDNER & DORLING, 1998).

An MLP consists of an input layer and an output layer with at least one additional hidden layer. Each layer can have multiple nodes and contains one or more processing units. Each unit in the MLP is completely interconnected with weighted connections (w_{ij}) to the units in the back layer. The output (o_j) is obtained by passing the sum of the inputs and weights by an activation function. Figure 4 shows a schematic of a simple neural network

MLP (GANDOMI & ROKE, 2015; VAPNIK & CORTES, 1995; ZHANG & MITTAL, 2013).

Figure 4-ANN general scheme of type MLP



Source: Adapted from Gandomi & Roke (2015)

The objective is to find an unknown function *f* which relates the input vectors in *x* to the output vectors in *y* by $Y = f(x)$, where X= [n x k] , Y = [n x j], n is number of training patterns, k is the number of input nodes/variables and j the number of output nodes/variables. During training, the function f is optimized, such that the network output for the input vectors in X is as close as possible to the target values in Y. The matrices X and Y represent the training data. The function *f*, for a given network architecture, is determined by the adjustable network weights. Both function approximation and prediction are very similar. To use an MLP for prediction involves training the network to output the future value of a variable, given an input vector containing earlier observations. The MLP approximates highly nonlinear functions between X and Y and requires no prior knowledge of the nature of this relationship. This is one of the benefits multilayer perceptrons offer over conventional regression analysis (GARDNER & DORLING, 1998).

2.4.3 Gaussian Process Regression

In recent years, Gaussian process regression (GPR) has received significant attention as a powerful statistical tool for data-driven modelling, it has been increasingly viewed as an alternative approach to ANNs, for example in applications such as spectroscopic calibration and nonlinear dynamic process modeling. The popularity of GPR is partly due to its theoretical link to Bayesian non-parametric statistics, infinite neural networks, kernel methods in machine learning, and spatial statistics. In addition, various empirical studies have

demonstrated that GPR attains prediction accuracy that is at least as comparable to (and in many cases better than) other models such as ANNs (NI *et al.*, 2011; WANG & CHEN, 2015).

The soft-sensor model development based on the GPR framework can be described as a problem whose aim is to learn a model f that approximates a training set $\{X\} = \{x_i\}_{i=1}^{N}$ and $\{Y\} = \{y_i\}_{i=1}^{N}$ are the input and output datasets with N samples, respectively. A GPR model provides a prediction of the output variable for an input sample through Bayesian inference. For an output variable of $y = [y_1, y_2, .., y_n]^T$, the GPR model is the regression function with a Gaussian prior distribution and zero mean, or in a discrete form (LIU & GAO, 2015)

$$y = (y_1, y_2, .., y_n)^T \sim G(0, C), \tag{29}$$

where C is the N x N covariance matrix with *ij*th element defined by the covariance function $C_{ij} = C(x_i, x_j)$. A common covariance widely used in the literature is:

$$C(x_i, x_j; \theta) = a_0 + a_1 \sum_{d=1}^{p} x_{id} x_{jd} + \nu_0 \exp\left(-\sum_{d=1}^{p} \eta_d (x_{id} - x_{jd})^2 \right) + \delta_{ij} b \tag{30}$$

where $\theta = (a_0, a_1, \eta_d, \ldots, \eta_p, \nu_0, b)$ is the vector of hyper-parameters, x_{id} is the *d*th component of the vector x_i. δ_{ij}=1 if i=j; otherwise it is equal to zero. The first two terms denote a constant bias and a linear correlation term respectively. The exponential term is similar to the form of RBF, recognizing high correlation between outputs with nearby inputs. In addition, the term b captures the random error effect. By combining both linear and nonlinear terms in the covariance function, the GPR model is capable of handling both linear and nonlinear processes (LIU & GAO, 2015; WANG & CHEN, 2015; NI *et al.*, 2011).

The hyper-parameters can be estimated by maximization of the corresponding log-likelihood function (Equation 31):

$$L(\theta) = -\frac{1}{2}\log(\det(C)) - \frac{1}{2} Y^T C^{-1} Y - \frac{N}{2}\log(2\pi) \tag{31}$$

After solving the optimization for θ, for a test sample x_q the predicted output of y_q is also Gaussian with mean

$$\widehat{y_q} = k_q^T C^{-1} Y \tag{32}$$

and variance

$$\sigma^2_{\widehat{y_q}} = r_q - k_q^T C^{-1} k_q \tag{33}$$

where k_q is the covariance between the new input and the training samples, r_q is the covariance of the new input and the vector $k_q^T C^{-1}$ denotes a smoothing term that weights the training outputs to make prediction for a new input x_q:

$$k_q = [C(x_q, x_1), C(x_q, x_2), \dots, C(x_q, x_N)]^T \quad (34)$$

$$r_q = C(x_q, x_q) \quad (35)$$

2.5 STATISTICAL PROCESS CONTROL (SPC)

Process always must be monitored, to detect special causes that increase dispersion or change in the mean from the target value. The main tool to monitor process and sign the presence of these causes are the control charts. There are different kinds of control charts and the adequate choice of one of them depends on the process and on the data to be monitored (RAMOS, 2012; COSTA *et al.*, 2005).

A phenomenon of autocorrelation has been found in many processes. When it happens, the control limits obtained by the conventional control charts cannot be applied without some changes; otherwise, the frequency of false alarm increases (RAMOS, 2000; COSTA *et al.*, 2005).

It is possible to use the mean and amplitude chart, or the individual chart. As this work used individual measurements of food quality variables, the second type of chart was applied with some necessary adjustments due to autocorrelation. The upper limit (ULC), the central limit (CL) and the lower limit (LLC) of control are calculated according to the followed Equations (36 - 39) (RAMOS, 2000; COSTA *et al.*, 2005):

$$\mathrm{ULC} = \bar{x} + 3\frac{s}{c_4} \quad (36)$$

$$\mathrm{CL} = \bar{x} \quad (37)$$

$$\mathrm{LLC} = \bar{x} - 3\frac{s}{c_4} \quad (38)$$

where s is the standard deviation calculated from the individual values and with the number of samples (k), by the equation above:

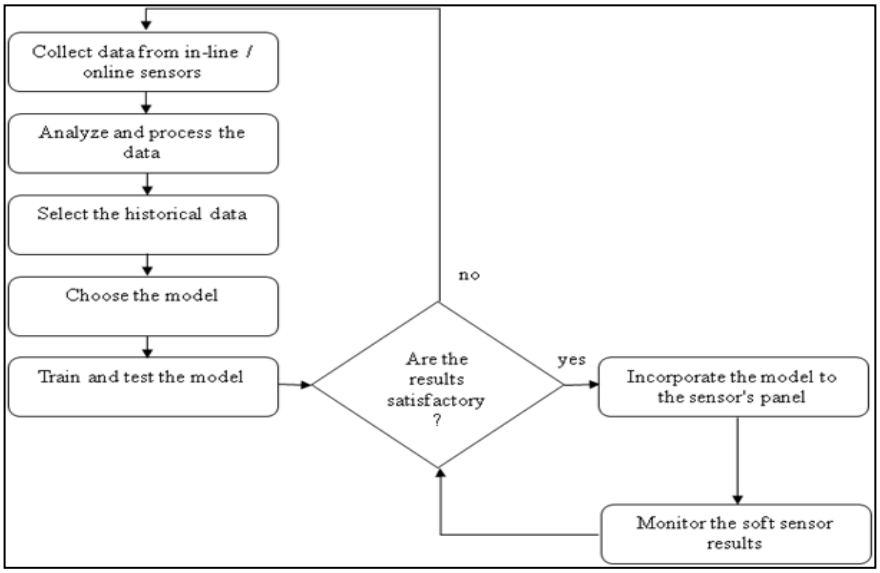
$$s = \sqrt{\sum \frac{(x_i - \bar{x})^2}{k - 1}} \quad (39)$$

And c_4 is the correction factor, when k>10, it is approximately 1.

3 BUILDING SOFT SENSORS TO MEASURE WATER ACTIVITY

In this chapter, the methodology for the elaboration and maintenance of SS to obtain a_w is proposed. It is based on the soft sensor methodology presented in KADLEC *et al.*, 2009. and also in the knowledge of machine learning, once it will be the soft sensors' base. The main steps of the a_w SS are presented in the flowchart of Figure 5.

Figure 5-Flowchart of the proposed methodology to build an a_w soft sensor



Source: The author (2019)

The first step to develop the a_w soft sensor is the data gathering, the user needs to evaluate which variables can be easily measured in the process, what sensors are available, evaluate if the variables have direct or indirect relation with a_w. Variables as temperature, pressure, moisture content, NaCl content, sugar content, brute protein, crude fiber, among others, can influence the interaction between water and the food and interfere in the a_w measurement. All these variables can be used as input to build the model of an a_w soft sensor.

The second step consists in the analysis of the obtained data set and to identify any obvious problem that could be handled at this initial stage. It is important to verify if they have the same measurement frequency, if there is the presence of outliers and deal with them, eliminate inconsistent values, as negative numbers, or values out of the possible range. Statistical tests are recommended to verify the correlation of these data with a_w, check if their means are the same or have significant differences (by the Kruskal-Wallis test, for example),

once this can interfere in the model performance. After evaluating the available data, and fixing any inconsistence, the data that will be actually used must be selected.

The next step is associated with the selection of the possible models that can be used to build the SS. The data set obtained in the previous step must be split into training and test sets. Depending on the model, it will be necessary to standardize or normalize the data. The standardization process just transforms the data to center by removing the mean value of each feature, then scale it by dividing non-constant features by their standard deviation. And the normalization is the process of scaling individual samples to have unit norm (SCIKIT LEARN, 2018).

Different models should be built and trained with the same training data set, and after that the test data set must be used to calculate error metrics such as Mean Absolute Percentage Error (Equation 40), Mean Absolute Error (Equation 41) and Mean Squared Error (Equation 42):

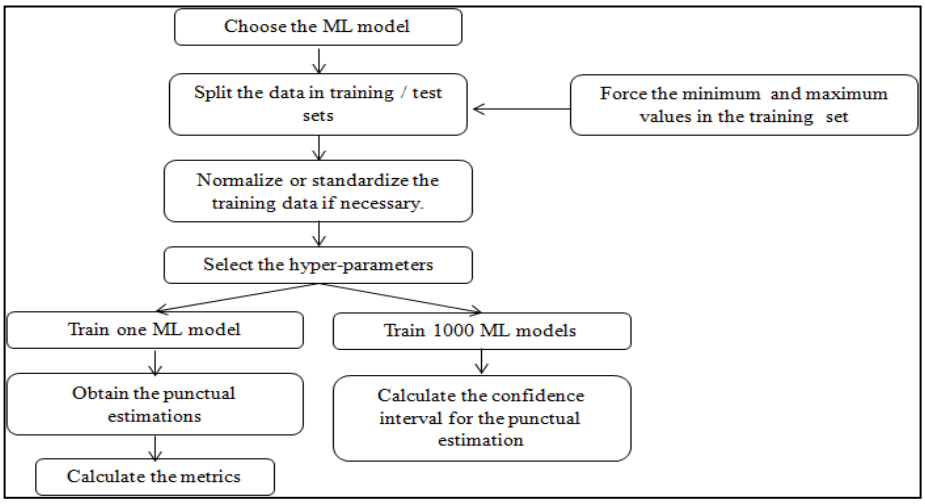
$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_{i,real} - y_{i,pred}}{y_{i,real}}\right| \tag{40}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{i,real} - y_{i,pred}\right| \tag{41}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{i,real} - y_{i,pred}\right)^2 \tag{42}$$

The results indicate which of them would be the best to be chosen as model for the soft sensor. If the metrics results are not satisfactory it's necessary change hyper-parameters, change the data, and repeat the training process. As regression models only provides point estimates related to the response variable of interest, bootstrap methods have become quite popular as tool for constructing confidence intervals within the framework of predictions via ML. It has a non-parametric characteristic that allows the construction of precise intervals without requiring assumptions about the data set underlying distribution (DAVISON; HINKLEY; YOUNG, 2003).

Figure 6 - Flowchart of the steps to train the ML model



Source: The author (2019)

After choosing the model, the estimated regression function must be incorporated to the sensor control panel. Thus, whenever the sensor measures the input variables, the panel can, in real-time calculate and provide the corresponding a_w value.

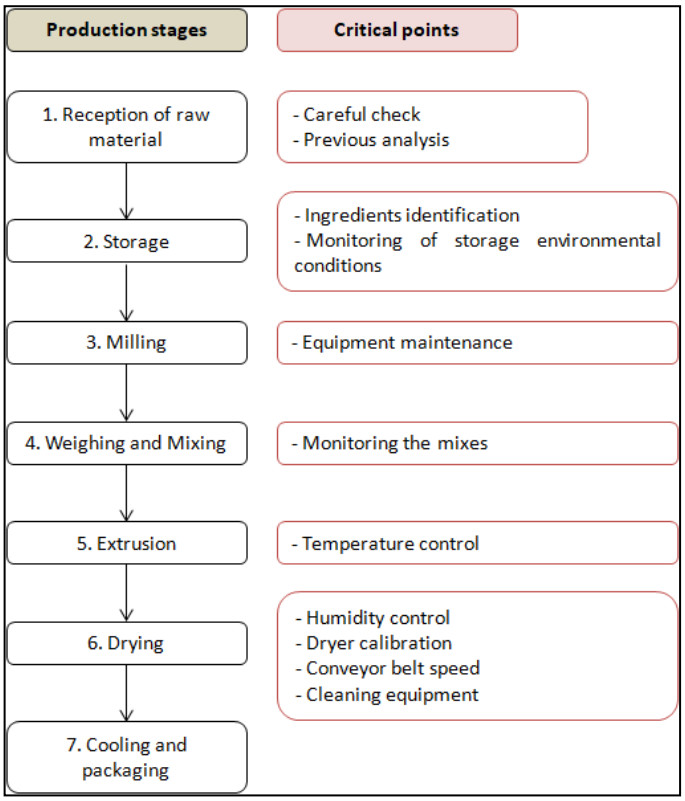
Once the model is incorporated to panel, it is necessary to care about its maintenance, due to drifts and/or other changes of data that may deteriorate its performance. When it occurs, a model adaptation or retraining using new gathered data is recommended, which is equivalent a calibration. For that, human intervention is still important through a periodic model checking, parallel analyses to validate a_w predicted values. This is necessary given that the model is generally valid for a specific situation, or process condition, which means that process modifications and variations may change the relation between a_w and the considered input variables.

A statistical process control can be used to help in this monitoring, once they are mainly applied to identify process changes, whether it is getting out of control. The input variables must be verified once they are used in the learning process to estimate a_w. If any change occurs, the model will continue to provide a_w estimates if no action is taken. In this context, statistical control charts can be used to monitor both a_w and the input variables, as they can easily indicate variation in the controlled variables due to possible changes in the production process. The best control chart will depend on the data and the process, whether there they present autocorrelation, for example.

4 APPLICATION EXAMPLE

It is notable the importance of food quality to human and animal consumption, considering the close relationship between them and their health. Thus, the control of critical factors related to pet food quality is necessary along all productive process. Pet food production process has several different stages, since the raw material (RM) reception until the product storage and distribution, Figure 7 (PEREIRA *et al.*, 2010)..

Figure 7- Flow-chart of pet food production stages and their critical points



Source: Adapted from Pereira *et al.*, 2010

Each stage has different critical points that need attention. The first step is the reception of raw material, where is necessary a careful check, because once unloaded it will be harder to separate the good from the low-quality material. A worker should anticipate himself doing analysis with the purpose of not acquiring low-quality RM. This stage aims at checking the presence of mycotoxins and other harmful microorganisms that may reduce food quality. For this reason, this phase must be done using the right procedures.

In the second stage, storage, it is important to avoid mix of ingredients or ingredients of different qualities. Therefore, they should be carefully identified. The storage time, the temperature, the relative humidity, the presence of animals that may contaminate products need to be monitored (FORMIGONI *et al.*, 2017; PEREIRA *et al.*, 2010).

The third stage consists in milling, where some of the products are evenly ground to be accepted by the animals. Also, equipment maintenance (mills) is necessary, with product moisture content, and the sieves' adjustments (rotation, distance). The weighing and mixing are a fundamental stage where the dosage and combination of ingredients occur to obtain great zoo-technical performances. It is essential to periodically monitor the mixes to guarantee a uniform product. In the extrusive stage, the paste acquires form, the ingredients after mixed passes through a mold or matrix. The paste also passes through processes involving high temperatures, which can interfere in the final product quality. Drying is the next stage and moisture control, dryer calibration, conveyor belt speed and cleaning equipment are some of the critical factors that contribute to make neither the paste have a superficial drying nor it dries so much. The last stage is cooling and packaging. Cooling serves to bring the food for the environmental temperature. After that, the packaging is done in a way to facilitate its transport (FORMIGONI *et al.*, 2017).

4.1 WATER ACTIVITY SOFT SENSORS FOR A PET FOOD PRODUCTION PROCESS

The application example presents different ML models to predict water activity from moisture content of a compound food: pet food. All measures of moisture content and a_w were provided by a pet food company. Based on this context, the proposed methodology to build a soft sensor was applied to this specific case. The soft sensor proposed is a real-time in-line soft sensor, once the time from the moisture content reading and the respective a_w computation doesn't exceed the residence time, and all of it is made direct in the process line.

Besides UM vs. a_w, the relation of a_w with other quality variables was observed. They are: brute protein (BP), mineral materials (MM), ethereal extract (EE), and crude fiber (FB) All quality variables measurements for a given sample of pet food were provided by a Near Infrared Spectrometer (NIR), except a_w value, which were observed in the same sample using a Pawkit.

This work was developed using Python version 3.6 and machine learning libraries: Scikit Learn library and package liquidSVM. All work was performed on a PC with an

Intel(R) Core (TM) i5-3337U CPU @ 1.80GHz, 6 GB RAM and Windows 10 operating system. The following steps were executed, considering the case of NIR, Pawkit, and a humidity sensor as the available equipment:

- Data inspection: the aim is to gain an overview of the data structure and to identify any obvious problems which may be handled at this initial stage. It was built a correlation heat-map and a scatter plot to visualize the correlation between each pair of variables. Statistical tests were performed to verify if the data came from the same distribution, once this can interfere in the model performance. It was evaluated if the presence of dye, different shifts (day and night) significantly interferes, as the environment temperature may impact food moisture content and a_w.
- Selection of historical data and identification of stationary states: the data used for the training and evaluation of the model were selected. In the case of batch processes there are usually no steady-states and thus the focus was on the selection of representative batch runs rather than on the identification of steady-states. An initial model was generated with data from a batch with the biggest range of moisture content. As it did not show a good result, probably due to the reduced size of the resulting data set. Thus, instead of considering a representative batch, all data was used and divided into training and test sets.
- Data preprocessing: the goal of this step is to transform the data so that it can be more effectively processed by the model. The main typical transformations are made by normalization and standardization. In this work, was verified the models' performance with and without the standardization of data. To evaluate the influence of this step, it was computed the score of the models using the test data set, and also it was verified their time consumption, where it was measured the time to find the best hyper-parameters and to train the model.
- Model selection, training and validation: the considered ML models were LR, SVR, LS-SVM, MLP and GPR. The kernel function of the SVM models and GPR was the Gaussian Radial Basis Function (RBF). A grid-search was applied to set the best hyper parameters for each model. The data set was randomly split in proportion of 70% and 30% for training and test, respectively. After training, the models' performances on test data were assessed by MAE, MSE and MAPE. Where the MAPE value of 10% was used

as reference for model acceptance (KAYMAR-ERTEKIN & GEDIK, 2004). Also, the time consumption during training phase was considerate.

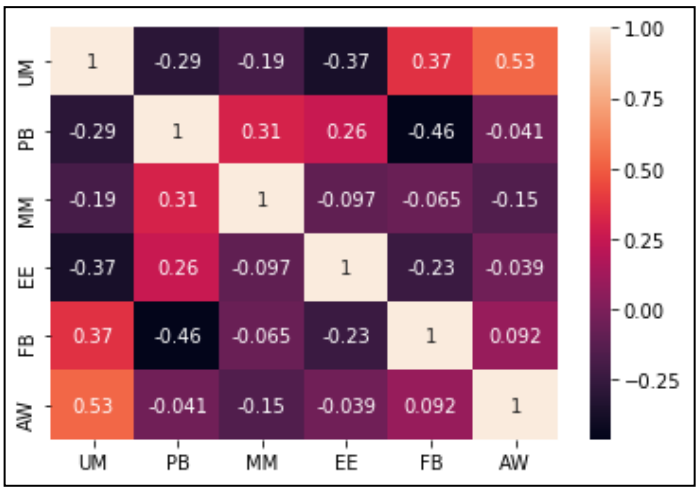
- Softer sensor maintenance: control charts that consider data autocorrelation were built for the moisture content measurements. proposed for this case take into account the existence of correlation between the data, what is present in the most case of batch process. Also, it was selected data from specific batch in order to evaluate the behavior of the best model in relation to these data, through the control charts.

4.1.1 Results

4.1.1.1 Data inspection and selection

This section presents a summary of the available data: measurements of moisture content (UM), brute protein (PB), mineral material (MM), ethereal extract (EE), brute fiber (FB) and water activity (a_w) were obtained for 252 pet food samples, using NIR and Pawkit, the latter used exclusively to provide a_w values. To assess the pairwise relation of these quality variables, the Pearson correlation coefficients were calculated. The heat-map of these quantities can be visualized in Figure 8.

Figure 8 - Heat-map of Spearman correlation between food quality variables

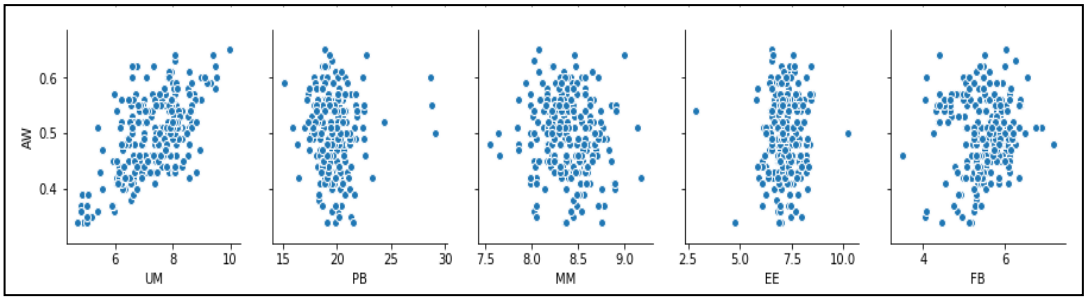


Source: The author (2019)

As can be seen in the heat-map, the moisture content presents a moderate positive linear correlation with water activity. These results are confirmed with the scatter plots of

Figure 9, which relates a_w with each of the input variables. The positive linear relationship between a_w and moisture content is evident from the corresponding scatter plot. The variables PB, EE, MM and FB did not present a similar behavior.

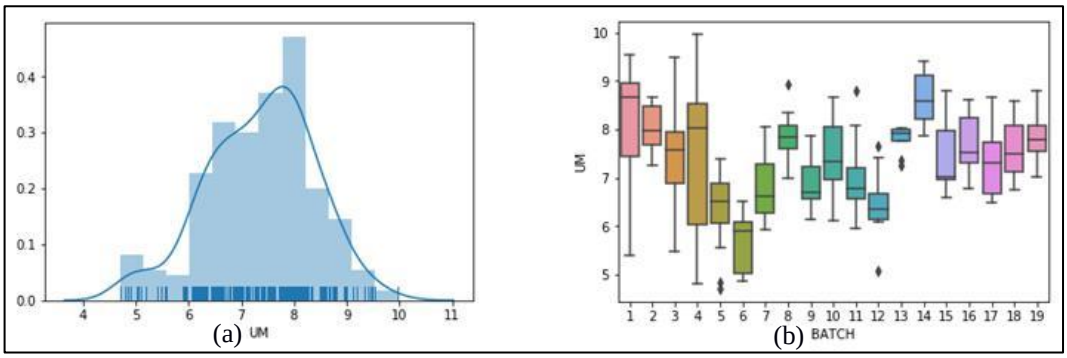
Figure 9- Scatter plot for the variable UM, PB, MM, EE and FB related with a_w



Source: The author (2019)

By considering the more evident relation of a_w and UM and given that the aim is at estimating a_w using data from a humidity sensor, the focus was on these two variables, having UM data as input and a_w as output. Figure 10 brings more details about UM data, its distribution and variability.

Figure 10- KDE plot and box-plot for UM data

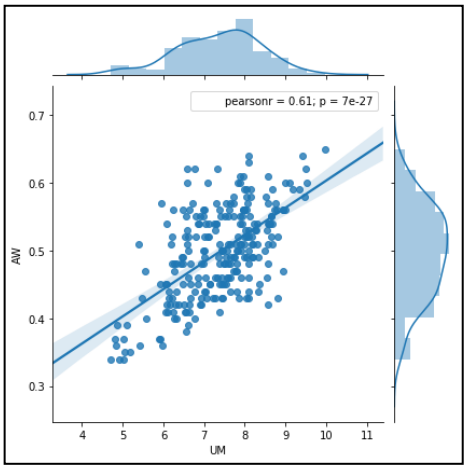


Source: The author (2019)

In Figure 10 (a) there is a distribution shape, like the histogram, the Kernel Density Estimator (KDE), which plots the density of observations on one axis with height along the other axis. This plot is a non-parametric way to estimate the probability function of a random variable. It is also possible to visualize that most of the data are centered between 6% and 9% of UM. They did not follow a normal distribution what was confirmed by statistical tests. In batch production, the phenomenon of variation is often present, that is, batches should be

homogeneous but there are some reasonable differences between them. This can be seen in the box-plot of Figure 10 (b). Figure 11 shows a scatter plot with a linear regression of UM and a_w, and also displays the KDE plot for each variable.

Figure 11- Linear regression plot using a_w and UM



Source: The author (2019)

By this first analysis the data should be selected removing the outliers, that we could visualize in the box-plot. In the case of batch processes, usually there is no steady-state and thus the model developer focuses on the selection of representative batch runs rather than on the identification of steady-states (KADLEC, *et al.* 2009).

Due to the small quantity of data collected, to the fact that there was no missing data and to the unidentified causes related to the outliers, all data was used to build the models presented.

4.1.1.2 Data pre-processing, model selection, training and validation

The LR, SVM, MLP, LS-SVM and GPR models were built following the steps presented before. The next figures present the plot of real data *versus* predicted data for all the models using the test dataset. Depending on the model, and the input data a standardization is a good pre-processing step. For our case the models presented better results without standardization, as can be seen in Table 2. It displays the time required for training, and the MAPE value obtained for all the models with and without standardization. Training time consumption is something important to care mainly when a large amount of data will be used

during training. The highlight in the case with standardization is for the LR model, which had the best training time consumption but it had the second worse MAPE training value (Table 2). Both values are important and trying to make a balance between them, the GPR presented the second training time and the second best MAPE training value, turning it a good option. For the case without standardization process the best MAPE training value came from the SVM model, but it had the longest training time. Concerning the balance between time consumption and MAPE training value the GPR presented the most promising results.

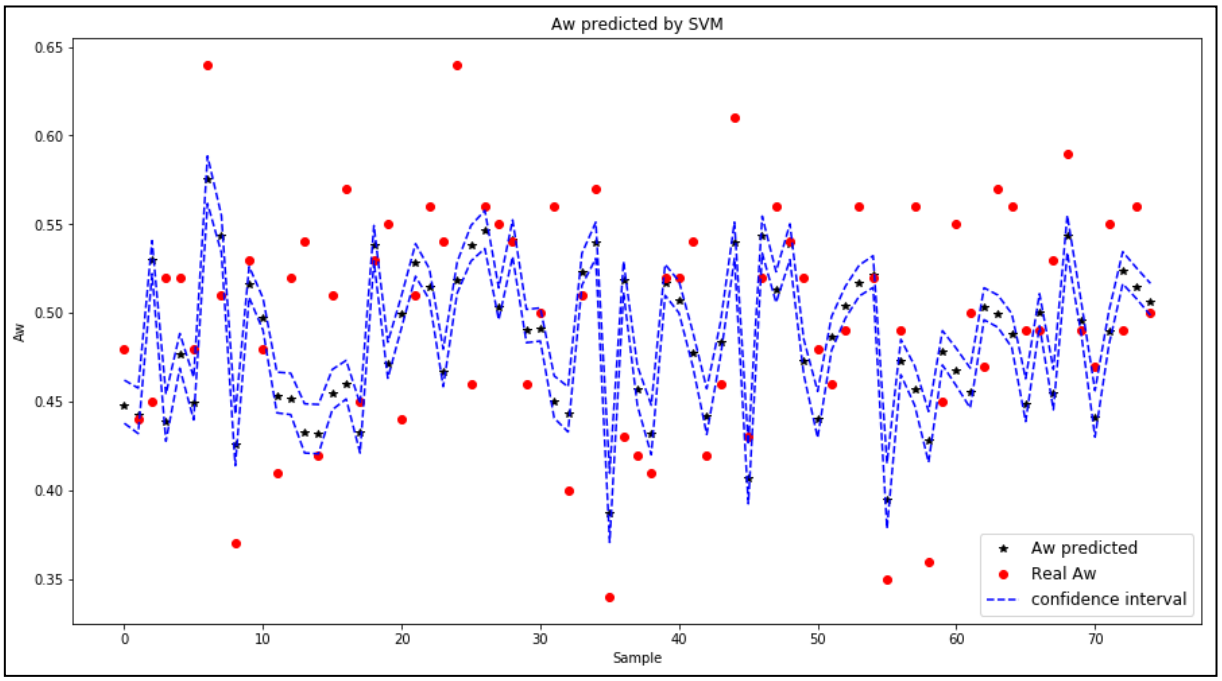
Table 2-Time consumption to train the ML models and the MAPE values calculated with training dataset before and after its standardization process

	Time (s)		**MAPE (%) training**	
Model	Without standard.	With standard.	Without standard.	With standard.
SVM	9.22	3.60	8.503	9.276
MLP	1.29	2.44	8.780	10.224
LSSVM	0.16	0.20	8.666	9.066
GPR	0.00	0.02	8.515	8.907
LR	0.00	0.00	8.779	10.069

Source: The author (2019)

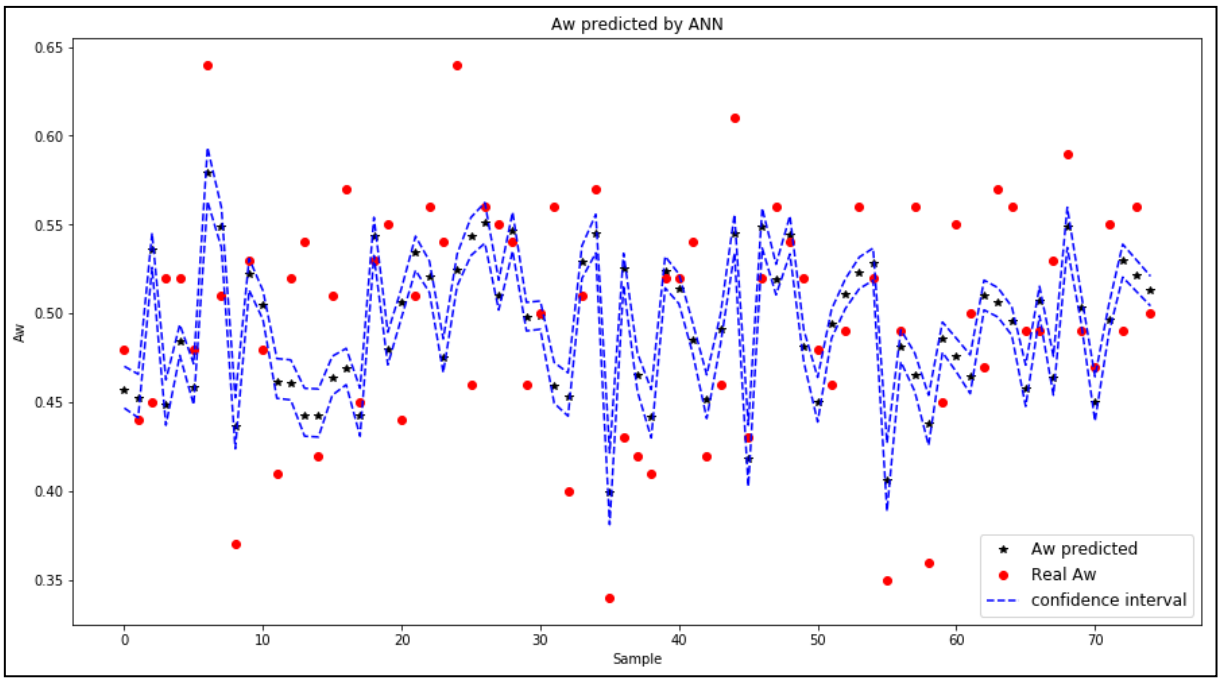
Figures 12-16 present the plots for all the test data set comparing the real a_w with the a_w predicted, for SVM, MLP, LS-SVM, GPR and LR models respectively, just for the case without standardization once most of the models presented the smaller training time and MAPE training value. The plot presents as dashed lines the confidence interval of 95% for these predictions, obtained by bootstrap method and using the percentile approach. The confidence interval defines the lower and upper limits that had, with some probability (confidence level), contain the true mean value. So that, if the confidence interval does not contain the true value, it should occur with probability of 5% or less. As it is possible to observe some points were out of the confidence interval, what probably occurs due to the non-repeatability of Pawkit measurements, which may cause large data variability.

Figure 12- Real a_w versus a_w predicted by SVM model



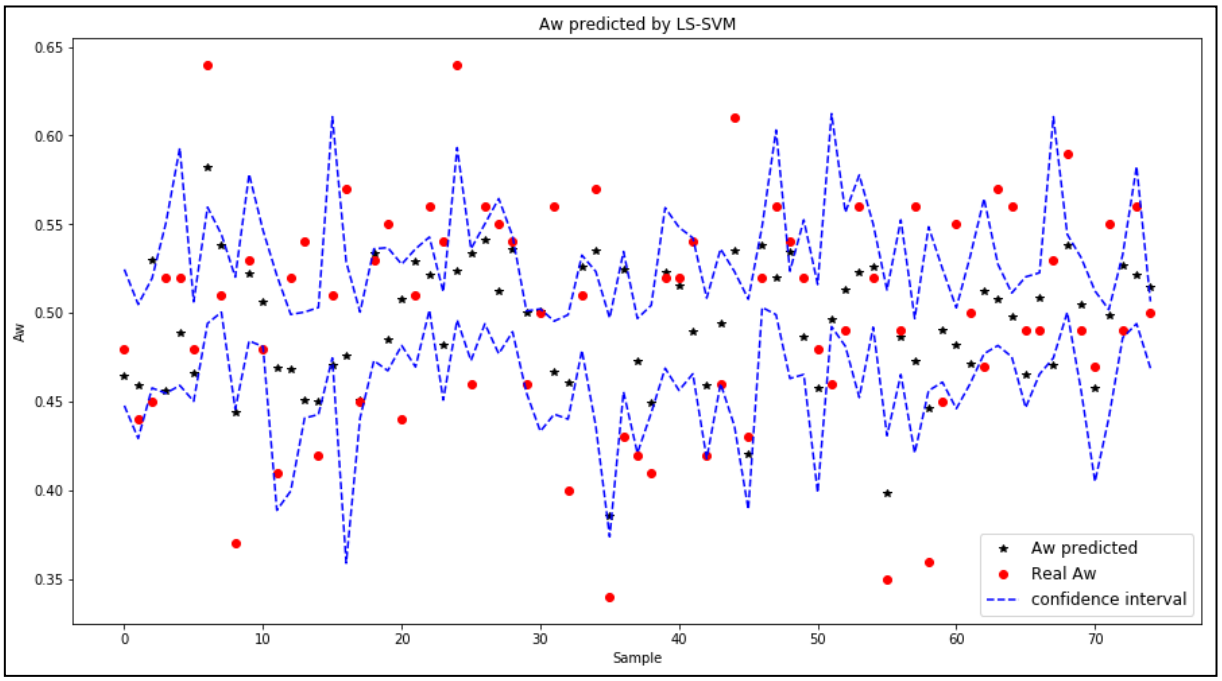
Source: The author (2019)

Figure 13- Real a_w versus a_w predicted by MLP model



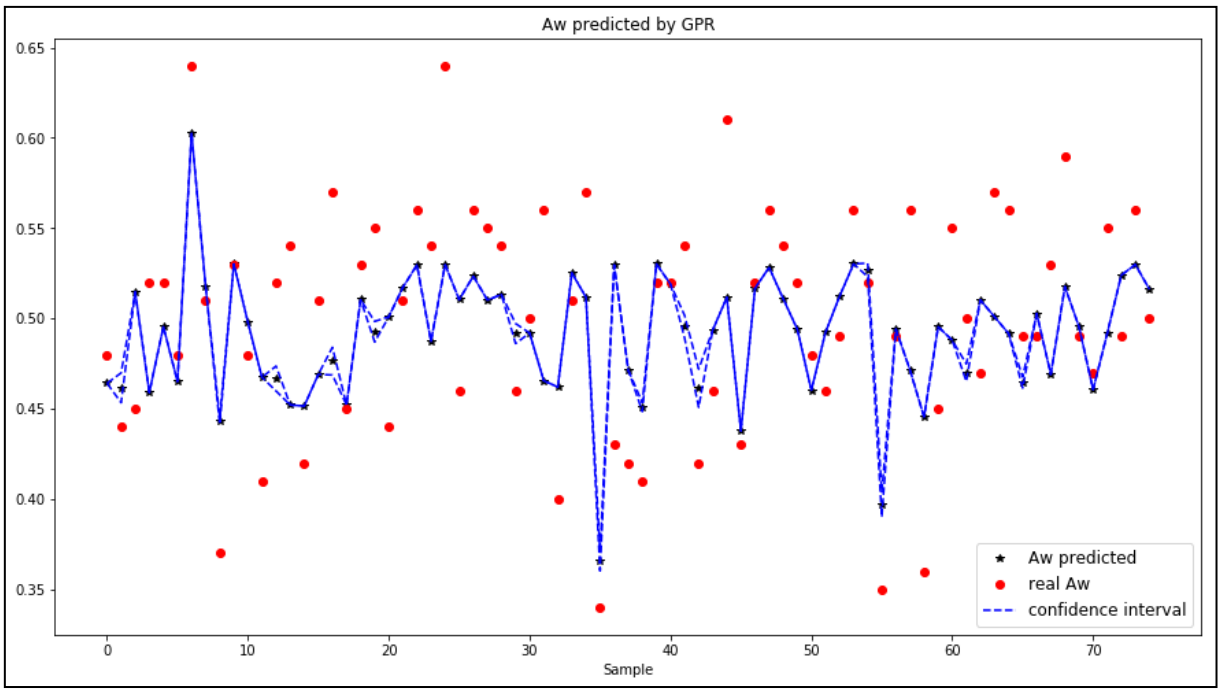
Source: The author (2019)

Figure 14- Real a_w versus a_w predicted by LS-SVM model



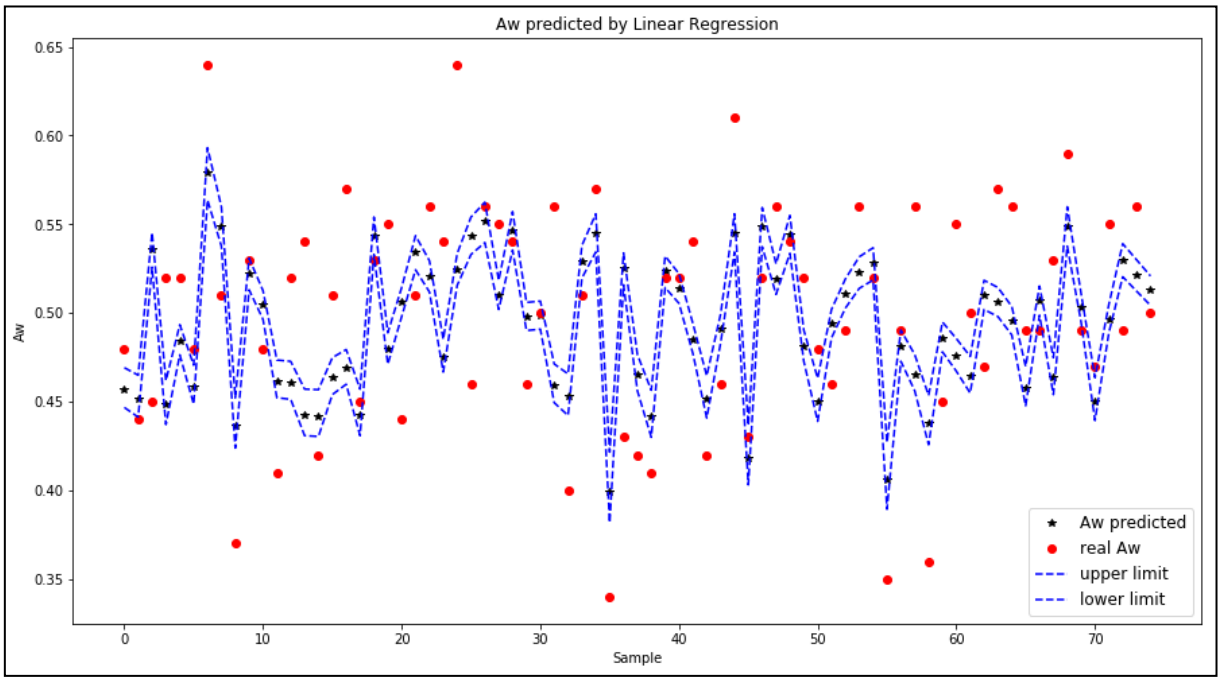
Source: The author (2019)

Figure 15- Real a_w versus a_w predicted by GPR model



Source: The author (2019)

Figure 16- Real a_w versus a_w predicted by LR model



Source: The author (2019)

Table 3 - Metric values of SVM, MLP, LS-SVM, GPR and LR models considering test data.

MODEL	MAPE(%) Test	MAE Test	MSE Test	Max. Absolute Error Test
SVM	8.575	0.043	0.0027	0.122
MLP	8.560	0.042	0.0025	0.115
LS-SVM	8.272	0.041	0.0024	0.116
GPR	8.051	0.040	0.0023	0.110
LR	8.561	0.042	0.0025	0.115

Source: The author (2019)

All the proposed models presented MAE and MSE values very close to each other approximately 0.04 and 0.003 respectively (Table 3), being necessary to evaluate other aspects before choosing the model that will be adopted in the soft sensor, as the MAPE Test value. The two best models according with the MAPE test values were the LS-SVM and GPR, however all the models could be considered acceptable once their MAPE test value was smaller than 10%. Considering the processing time as an important criterion to choose a model once it would feed a soft sensor that will use a large quantity of data, the GPR and LR models were faster than the others (Table 2). The GPR model also showed the best result when comparing the maximum error test, what means that between the actual and predicted

values the biggest absolute error found for the test data set was 0.110, which reinforces the option to chose GPR.

4.1.1.3 Testing with a data provided by a humidity sensor

After building the statistical model presented, an evaluation was proposed for the GPR model. The aim was at verifying their performance given a set of data obtained directly from a humidity sensor, considering that this sensor was calibrated by the same equipment NIR used to provide the input data that feed the ML model. For that, a_w was predicted for measures of moisture content obtained by humidity sensor, and a_w was compared with the value obtained by the Pawkit. Table 4 presents the metrics MAE, MAPE and the maximum differences found, for all five models using the data collected by the sensor.

Table 4-Models assessment and respective errors for each built model

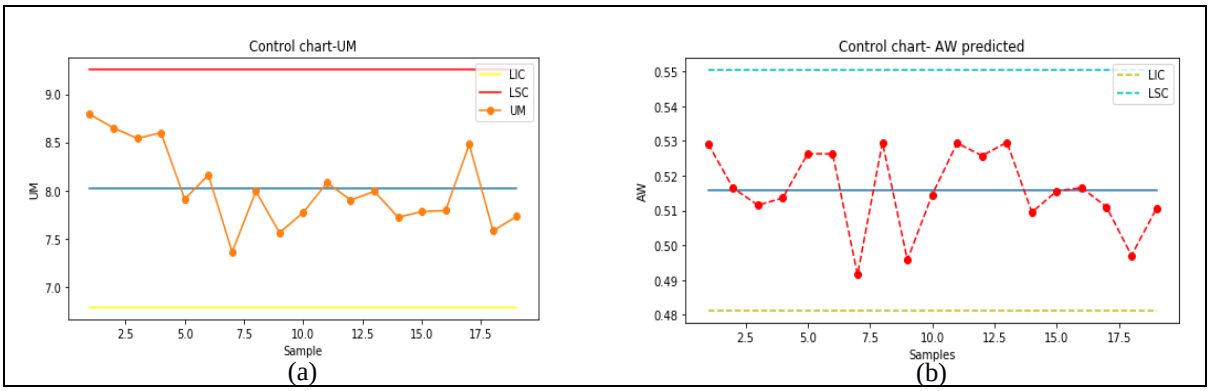
MODEL	MAE	MAPE(%)	Max. Absolute Error Test
SVM	0.037	7.892	0.071
MLP	0.036	7.690	0.070
LS-SVM	0.031	6.834	0.071
GPR	**0.025**	**5.675**	**0.051**
LR	0.036	7.694	0.067

Source: The author (2019)

4.1.1.4 Soft sensor maintenance

The models' performances depend on the training data, if the process condition changes, the model would not be able to predict a_w correctly, and model retraining would be necessary. To identify the process variations, control charts can be used to monitor the input data (UM) and the output data (estimated a_w). If the UM measures displayed a change in the process there was an indicator of necessity to verify the a_w results together with a laboratory, for example. The easiest way to verify this process drift is by a control chart. For our work and kind of process, it was proposed a control chart considering the existence of autocorrelation. Figures 17 presents control charts obtained for part of the data provided, to observe and compare the behavior of UM data and a_w predicted by the GPR model. The behaviors of predicted a_w and UM are similar, as expected.

Figure 17- (a)Control chart of UM. (b) a_w predicted for the GPR model.



Source: The author (2019)

After the calculation of these limits, they can be used in the panel connected to the humidity sensor to alert when the measured values exceed them. It is also important to keep a periodical checking and control to avoid any measurement problems and to know if the model is still operating as expected.

5 CONCLUSIONS

This work had as objective to propose a methodology to create a static soft sensor to predict water activity values from measures provided by available sensors as a humidity sensor, that already was working in-line/on-line. Five machine learning algorithms (LS-SVM, SVM, GPR, MLP, and LR) were trained and based the soft sensor methodology, were applied in an example involving data of a compound food provided by a pet food industry. Water activity values of 75 samples were predicted based in the entrance of the respective moisture content, the models' performances were measured by the training time consumption, metrics as MAPE, MAE, MSE values and the maximum absolute error obtained with the test data set. The GPR model presented the best performance. After that the models predicted the a_w value expected for moisture content measured by a humidity sensor, which confirm the best performance of the GPR model. However, considering a MAPE value smaller than 10% as reference to classify a model as acceptable, all of them could be considered satisfactory.

After model's choice to base the soft sensor, it is possible to incorporate it in the panel of the available sensor to predict a_w in real-time, avoid the necessity of sampling in a short frequency, and reduce the operator participation in the quality measurement process.

Among the limitations of these models, the necessity to collect a sample periodically and make laboratory analysis to verify if the model keeps performing well. The use of control charts was also suggested to monitor the soft sensor predictions. The kind of soft sensor proposed is based on empirical observation; thus, it describes the true conditions of the process and can be used in complex industrial processes. The portable equipment used have a precision of 0.02, which influences model precision. The only available equipment in-line/non-invasive was a humidity sensor that was installed and calibrated based on NIR.

Given that, there is a range of possibility of future works as test the models proposed with other available equipment, a NIR in-line for example. Also make a study with a large range of temperature, changing the sugar and salt content in the composition of the food in order to evaluate how it can affect the results and the models. With a large quantity of data, the use of other type of algorithms suited for deep learning becomes possible, for example convolutional neural networks. There is also the possibility of process integration in the sense that the output of the soft sensor can be used to trigger automated adjustments in the production process.

REFERENCES

ACHATA, E.; ESQUERRE,C.; O'DONNELL, C.; GOWEN,A. A study on the application of near infrared hyperspectral chemical imaging for monitoring moisture content and water activity in low moisture systems. *Molecules*, v. 20, n. 2, p. 2611–2621, 2015.

ACOSTA, S. M.; AMOROSO, A.L; SANTANNA, A.M.O. Modelagem de um processo produtivo utilizando regressão por vetores suporte. In XXXVI Encontro Nacional de Engenharia de Produção, 36. , 2016, João Pessoa. *Anais*[...] João Pessoa: ABEPRO, 2016.

ALFAMARE, *Medidor de atividade de água portátil - Pawkit*; Available in: https://alfamare.com.br/produtos/medidor-de-atividade-de-agua-portatil-pawkit/. Accessed in: 28 September 2018.

ALPAYDIN, E. *Introduction to machine learning*. 2 Ed. Cambridge: The MIT Press, 2010.

AWAD, M.; KHANNA, R. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designe.* Apress Open, 2015.

BALABIN, R. M.; LOMAKINA, E. I. Support vector machine regression (SVR / LS-SVM)— an alternative to neural networks (ANN) for analytical chemistry ? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. *Analyst*, v. 136, n. 8, p. 1703–1712, 2011.

BAPTISTA, D.; MORGADO-DIAS, F. A survey of artificial neural network training tools. *Neural Computing and Applications*, v. 23, n. 3–4, p. 609–615, 2013.

BAŞER, Ö.; YALÇIN, S. Determination of some quality characteristics in pet foods. *Ankara Üniv Vet Fak Derg*, v. 64, n. 1, p. 21, 2017.

BRUWER, M.J.; MACGREGOR,J.F.; BOURG JR, W.M. Soft Sensor for Snack Food Textural Properties Using On-Line Vibrational Measurements. *Industrial & Engineering Chemistry Research*, v.46, n. 3, p. 864-870, 2007.

BURGES, C. J. C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, v. 167, p. 121–167, 1998.

CAZIER, J.P., GEKAS, V. Water activity and its prediction: a review water activity and its prediction. *International Journal of Food Properties*, v. 2912, n. 2001, 2007.

CESAR, J.; ALVES, L.; POPPI, R. J. Talanta Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM). *Talanta*, v. 104, p. 155–161, 2013.

CHAURASIA, P. *et al.* Comparison of Gaussian process regression , artificial neural network, and response surface methodology modeling approaches for predicting drying time of mosambi (Citrus limetta) peel. n. November, p. 1–8, 2018.

CIFUENTES, A. Food Analysis: Present, Future, and Foodomics. *ISRN Analytical Chemistry*,

v. 2012, p. 1–16, 2012.

Collell,C.; Gou, P.;Picouet, P.; Arnau, J.; Comaposada, J. Feasibility of near-infrared spectroscopy to predict aw and moisture and NaCl contents of fermented pork sausages. *Meat Science*, v.85, n.2, p.325-330

COSTA, A.F.B; EPPRECHT, E.K.; CARPINETTI, L.C.R. *Controle estatístico de qualidad,* 2. ed. São Paulo: Atlas, 2005.

DAVISON, A. C.; HINKLEY, D. V; YOUNG, G. A. Recent Developments in Bootstrap Methodology. *Statistical Science*, v. 18, n. 2, p. 141–157, 2003.

DECAGON DEVICES. *AquaLab Pre*, 2012. Available in: http://aqualab.decagon.com.br/assets/Uploads/13484-08-AquaLab-4.pdf. Acessed in: 05 May 2018

ESCOBAR, C. A.; MORALES-MENENDEZ, R. Machine learning techniques for quality control in high conformance manufacturing environment. *Advances in Mechanical Engineering*, v. 10, n. 2, p. 1–16, 2018.

FDA, *Water activity (aw) in foods*. 2015 Available in: https://www.fda.gov/iceci/inspections/inspectionguides/inspectiontechnicalguides/ucm072916.htm. Accessed in: 15 March 2018.

FERBLANTIER, G.; SORLI, B.; PASCAL-DELANNOY, F.; BOYER, A. Feasibility of a new sensor for water activity determination. *Sensor Review*, v. 23, n. 4, p. 316–322, 2003.

FERRÃO, M. F.; MELLO, C.; BORIN,A.; MARETTO, D.A.; POPPI, R.J. LS-SVM: Uma nova ferramenta quimiométrica para regressão multivariada. Comparação de modelos de regressão LS-SVM e PLS na quantificação de adulterantes em leite em pó empregando NIR. *Quimica Nova*, v. 30, n. 4, p. 852–859, 2007.

FORMIGONI, A. DA S.; MARCELO, G. DE C.; NUNES, A. N. Importância do programa de qualidade “boas práticas de fabricação” (BPF) na produção de ração. *Nutritime Revista Eletrônica*, v. 14, n. 6, p. 10, 2017.

FRANKS, F. Water activity: a credible measure of food safety and quality? *Trends in Food Science and Technology*, v. 2, n. C, p. 68–72, 1991.

GANDOMI, A. H.; ROKE, D. A. Assessment of artificial neural network and genetic programming as predictive tools. *Advances in Engineering Software*, v. 88, p. 63–72, 2015.

GARDNER, M. W.; DORLING, S. R. Artificial Neural Networks (The multilayer percerptron)— A review of applications in the atmospheric sciences. *Atmospheric Environment*, v. 32, n. 14, p. 2627–2636, 1998.

KADLEC, P.; GABRYS, B.; STRANDT, S. Data-driven Soft Sensors in the process industry. *Computers and Chemical Engineering*, v. 33, n.4, p. 795–814, 2009.

LEWICKI, P. P. Water as the determinant of food engineering properties. A review. *Journal of Food Engineering*, v. 61, n. 4, p. 483–495, 2004.

LI, H.; LIANG, Y.; XU, Q. Support vector machines and its applications in chemistry. *Chemometrics and Intelligent Laboratory Systems*, v. 95, n. 2, p. 188–198, 2009a.

LIU, G.; ZHOU, D.; XU, H.; MEI, C. Model optimization of SVM for a fermentation soft sensor. *Expert Systems With Applications*, v. 37, n. 4, p. 2708–2713, 2010.

LIU, Y.; GAO, Z. Industrial melt index prediction with the ensemble anti-outlier just-in-time Gaussian process regression modeling method. *Journal of Applied Polymer Science*, v. 41958, p. 1–10, 2015.

LORENA, A. C.; CARVALHO, A. C. P. L. F. DE. Uma Introdução às Support Vector Machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007.

LUO, G. A review of automatic selection methods for machine learning algorithms and hyper-parameter values. *Network Modeling Analysis in Health Informatics and Bioinformatics*, v. 5, n. 1, p. 1–16, 2016.

LUTS, J. *et al.* A tutorial on support vector machine-based methods for classification problems in chemometrics. *Analytica Chimica Acta*, v. 665, n. 2, p. 129–145, 2010.

MATHLOUTHI, M. Water content, water activity, water structure and the stability of foodstuffs. *Food Control*, v.12, n. 7, p. 409 - 417, 2001.

MCMEEKIN, T. A. *et al.* Quantitative Microbiology: A Basis for Food Safety. *Emerging Infectious Diseases*, v. 3, n. 4, p. 541–549, 1997.

METER GROUP, I. U. Aqualab series 4®. Available in: http://library.metergroup.com/Manuals/20115_AQUALAB4_Manual_Web.pdf. Accessed in: 19 October 2018.

MOTA DA SILVA, V.; SILVA, L.A.; ANDRADE, J.B.; VELOSSOL, M.C.C.; SANTOS, G.V. Determination of moisture content and water activity in algae and fish by thermoanalytical techniques. *Química Nova*, v. 31, n. 4, p. 901–905, 2008.

MUÑOZ, I. *et al.* Drying control of cured sausages through online measurement of product quality. *Drying Technology*, v. 25, n. 11, p. 1809–1817, 2007.

NI, W.; TAN, S. K.; NG, W. J. Recursive GPR for nonlinear dynamic process modeling. *Chemical Engineering Journal*, v. 173, n. 2, p. 636–643, 2011.

NIELSEN, J.; ANDREASEN, M.B.; PEDERSEN, M;, RASMUSSEN, M.K.;Towards Fast In-line Measurement of Water Activity. *International Journal of Thermophysics*,Volume 36, n. 2–3, p. 577–588, 2015.

OSORIO, D.; PÉREZ-CORREA, J.R.; AGOSIN, E.; CABRERA, M. Soft-sensor for on-line estimation of ethanol concentrations in wine stills. *Journal of Food Engineering*, v. 87, n.4, p.

571-577, 2008.

PARK, K.J.B.; PARK, K.J., CORNEJO, F.E.P., FABBRO, I.M.D.; Considerações termodinâmicas das isotermas. *Revista Brasileira de Produtos Agrícolas*, v.10, n.1, p. 83-94, 2008.

PEDERSEN, J. N. M. B. A. M.; RASMUSSEN, M. K. Towards Fast In-line Measurement of Water Activity. p. 577–588, 2015.

PEREIRA, A.; MACHADO, L. C.; NORONHA, M. S. Controle de qualidade na produção de rações. *PUBVET, Publicações em Medicina Veterinária e Zootecnia.*, v. 4, n. 134, p. 33, 2010.

PORTER, R.; THEILER, J.; HUSH, D. Interactive Machine Learning in Data Exploitation. *Computing in Science & Engineering*, v. 15, p. 12–20, 2013.

PRABHAKAR, K.; MALLIKA, E. N. Water Activity. *Encyclopedia of Food Microbiology*, v. 3, p. 751–754, 2014.

PRIOR, B. A. Measurement of Water Activity in Foods - A Review. *Journal of Food Protection*, v. 42, n. 18, p. 668–674, 1979.

RAMOS, A.W. *CEP para processos contínuos e em bateladas.* 1. ed. São Paulo: Fundação Vanzolini, Edgard Blücher, 2012.

ROOS, Y. H. Water activity - Principles and Measurement. *Elsevier Science*, 2003.

SABLANI, S.S. ; KASAPIS, S.; RAHMAN, M. S. Evaluating water activity and glass transition concepts for food stability. *Journal of Food Engineering*, v. 78, p. 266–271, 2007.

SCIKIT LEARN, Supervised Learning. Available in: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning. Accessed in: 28 September 2018.

SIMHA, H. V. V.; PUSHPADASS, H. A.; FRANKLIN, M. E. E.; KUMAR, P. A.; MANIMALA, K. Soft computing modelling of moisture sorption isotherms of milk-foxtail millet powder and determination of thermodynamic properties. *Journal of Food Science and Technology*, v. 53, n. 6, p. 2705–2714, 2016.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. *Statistics and Computing*, v. 14, n.3, p. 199–222, 2004.

SOUZA, F. A. A.; ARAÚJO, R.; MENDES, Review of soft sensor methods for regression applications. *J. Chemometrics and Intelligent Laboratory Systems*, v. 152, p. 69–79, 2016.

TIMMERMANN, E. O.; CHIRIFE, J.; IGLESIAS, H. A. Water sorption isotherms of foods and foodstuffs : BET or GAB parameters ? *Journal of Food Engineering*, v. 48, p. 19–31, 2001.

TRAFFANO-SCHIFFO, M.V.; CASTRO-GIRALDEZ, M.; COLOM, R.J.; FITO, P.J. Study

of the application of dielectric spectroscopy to predict the water activity of meat during drying process. *Journal of Food Engineering*, v. 166, p. 285-290, 2015.

TREVISAN, MARCELLO G; POPPI, R. J. Química Analítica de Processos. *Quimica Nova*, v. 29, n. 5, p. 1065–1071, 2006.

TROLLER, J. Methods to Measure Water Activity. *Journal of Food Protection*, v. 46, n. 2, p. 129–134, 1983.

TSENG, T. L.; ALETI, K.R.; HU, Z.; KWON, Y. E-quality control: A support vector machines approach. *Journal of Computational Design and Engineering*, v. 3, n. 2, p. 91–101, 2016.

VAPNIK, V.; CORTES, C. Support-vector Networks. v. 7, p. 273–297, 1995.

WANG, L. Support Vector Machines: Theory and Application. Springer, 2005.

WANG, B.; CHEN, T. Gaussian process regression with multiple response variables. *Chemometrics and Intelligent Laboratory Systems*, v. 142, p. 159–165, 2015.

WANG, H.; HU, D. Comparison of SVM and LS-SVM for regression. *IEEE Transactions on Industrial Electronics*, n. 5, p. 279–283, 2005.

YU, J. A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses. *Computers and Chemical Engineering*, v. 41, p. 134–144, 2012.

ZHANG, J.; MITTAL, G. S. Prediction of equilibrium moisture content of various foods using a neural network. *International Journal of Food Properties*, v. 16, n. 6, p. 1211–1221, 2013.