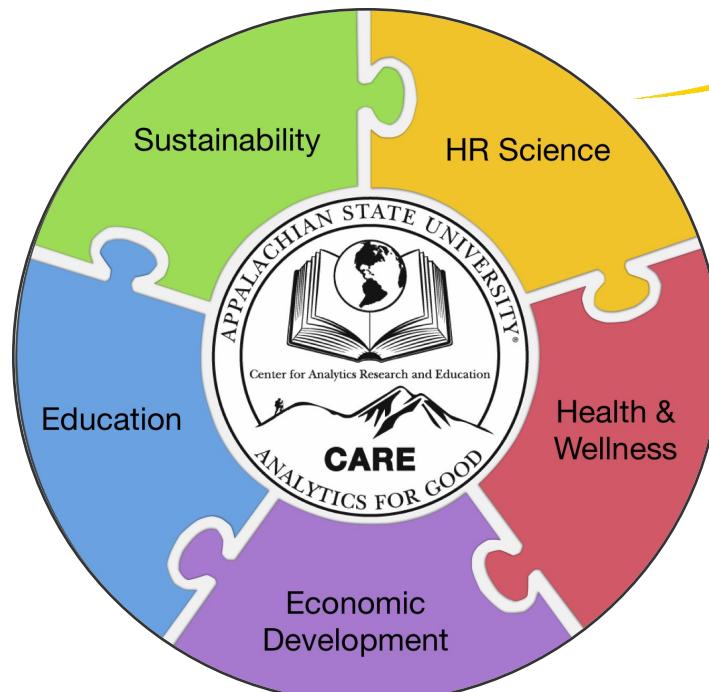


# A Cluster-Classification Method for Accurate Mining of Seasonal Honey Bees Colonies Patterns



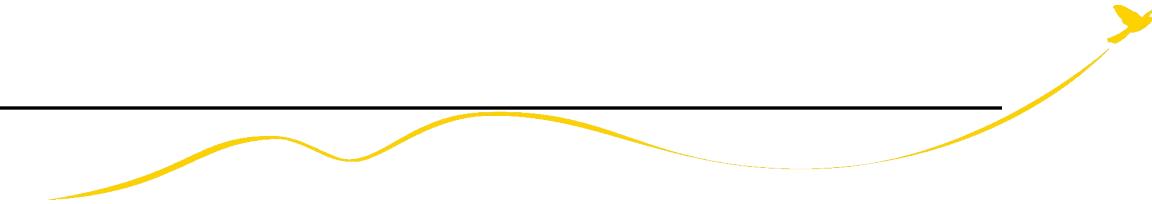
Antonio Rafael Braga, Ph.D Candidate

Center for Analytics Research and Education  
February 16, 2019  
SAHRC



# Agenda

---

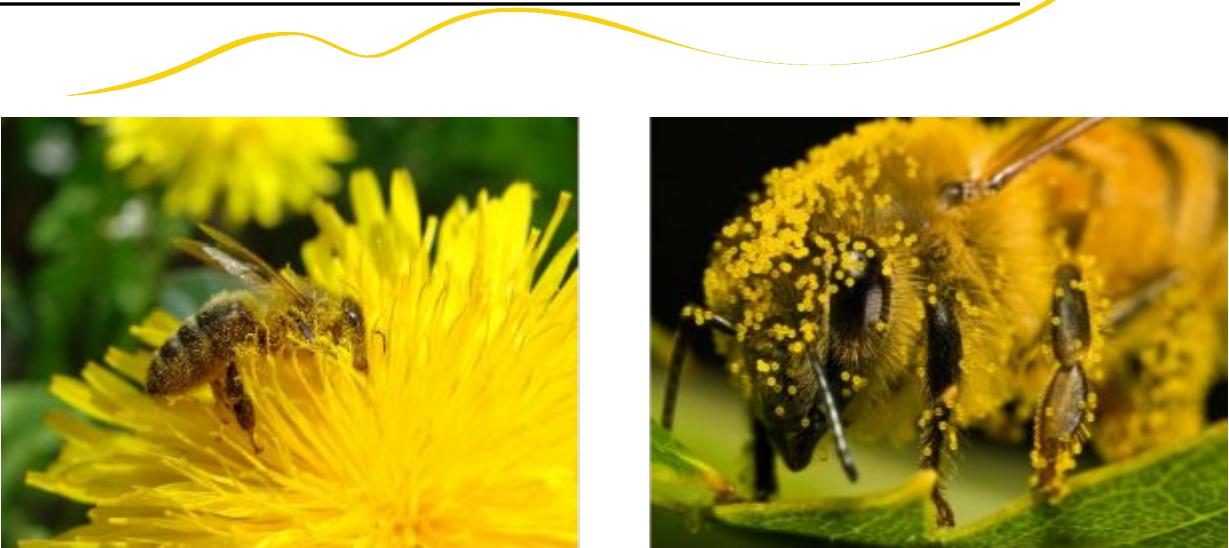


1. Introduction
  - Motivation
  - Contribution
2. Related works
3. Materials and methods
  - Datasets
  - Pre-Processing
  - Learning Strategies
4. Experimental Evaluation
5. Discussion
6. Conclusion

# Motivation

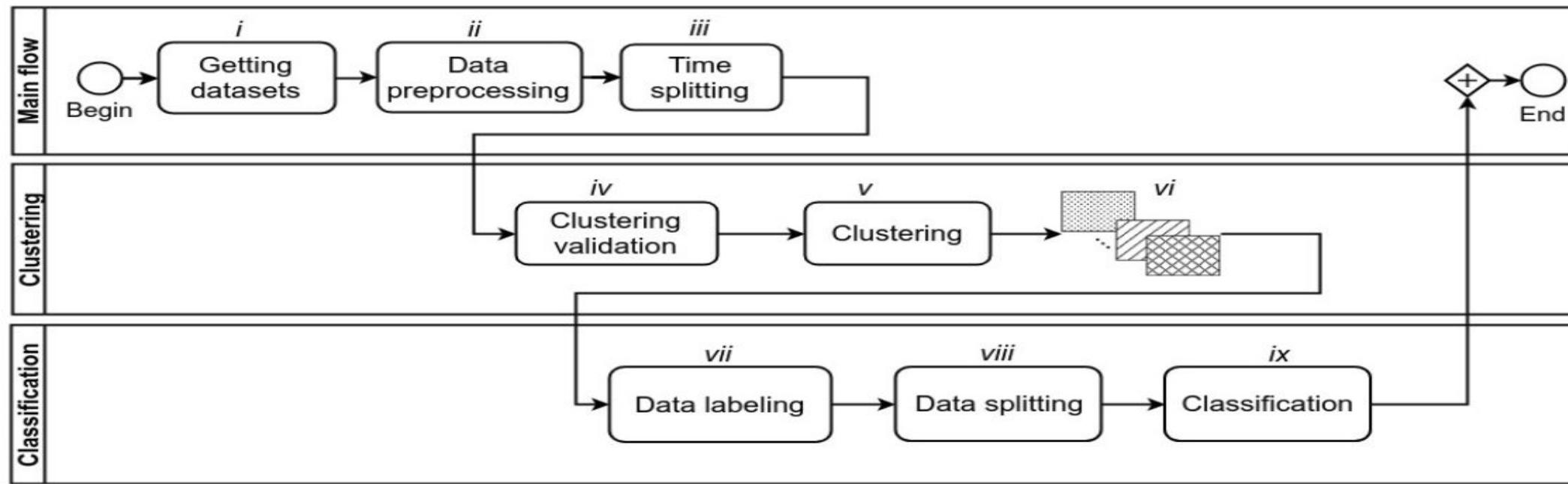
## Bees

- Main pollinating agents;
  - 75% of plant crops;
- Essential for food production;
- Maintenance of ecosystems;
- Different states observed during the life cycle of a colony (swarming, brood rearing, preparation for overwinter);
- Precision Beekeeping - remote monitoring of apiaries (WSN and IoT);
- Data semantics;

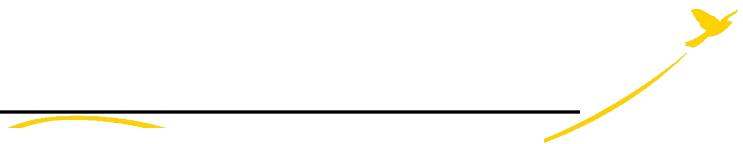


# Research questions (RQ)

- RQ#1: "What are the typical values of temperature, relative humidity and weight for a strong or weak colony on a yearly-basis and between the seasons?"
- RQ#2: "Is it possible to determine the health status of a beehive from homeostasis (brood temperature/relative humidity), productivity (hive weight) and external conditions (temperature/relative humidity) data?".



# Related works



Reference	Metrics	Technique
Zacepins and Karasha (2012)	Temperature	—
Zacepins et al. (2016)	Temperature	Decision support algorithm
Zacepins et al. (2017)	Temperature	—
Kridi et al. (2014)	Temperature	Clustering
Kridi et al. (2016)	Temperature	Clustering
Kviesis and Zacepins (2016)	Temperature	Neural networks
Fitzgerald et al. (2015)	Weight	—
Ruan et al. (2017)	Weight	—
Gil-Lebrero et al. (2017)	Temperature, humidity and weight	—
Murphy et al. (2016)	Temperature, humidity and gases	Decision trees
Murphy et al. (2015a)	Audio	Signal processing
Murphy et al. (2015b)	Audio, video and vibration	Signal and image processing
Tashakkori and Ghadiri (2015)	Video	Image processing
Chazette et al. (2016)	Temperature, weight, audio and video	Image processing
Bencsik et al. (2011)	Vibration	PCA
Bencsik et al. (2015)	Vibration	DFA
Chen et al. (2012)	Number of bees	SVM
Meikle et al. (2017)	Temperature and humidity	ANOVA

# Datasets

- **Portal:** HiveTool.net;
- **Beehives:**
  - Arnas (Denmark) – 92.749 samples;
  - Emil (Norway) – 88.181 samples;
- **Period:** 01/03/2017 a 28/02/2018;
- **Features:** Temperature (internal and external), humidity (internal and external) e weight;

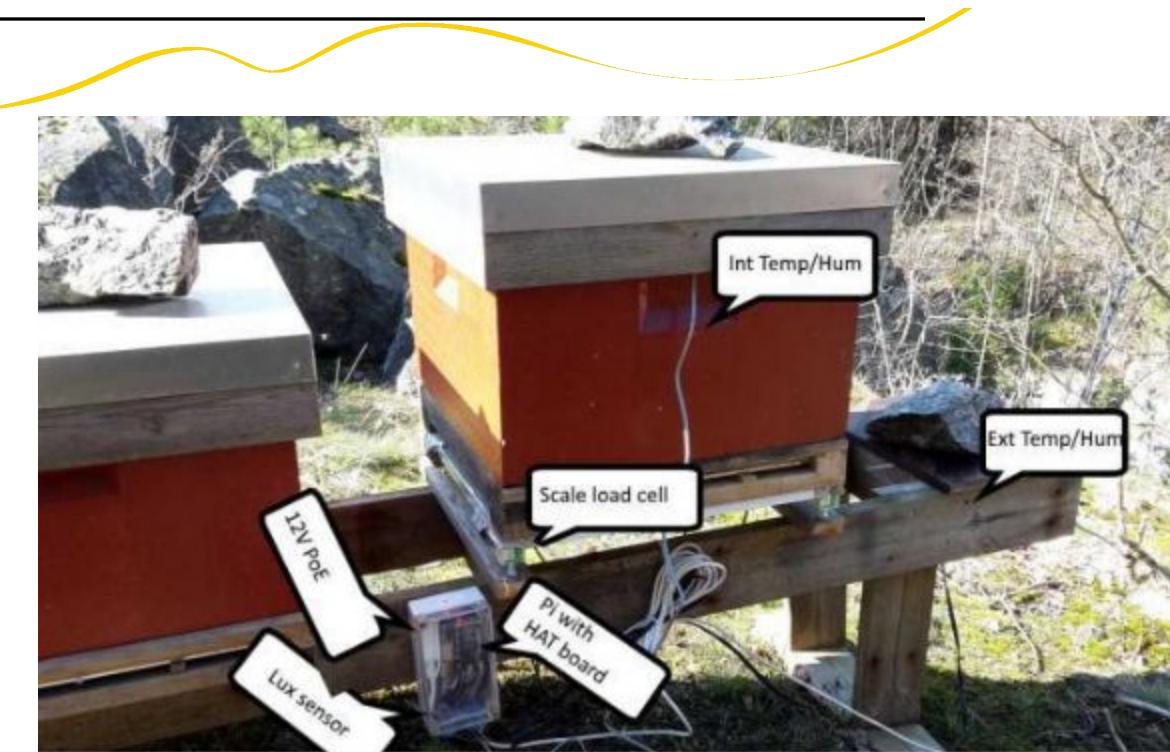
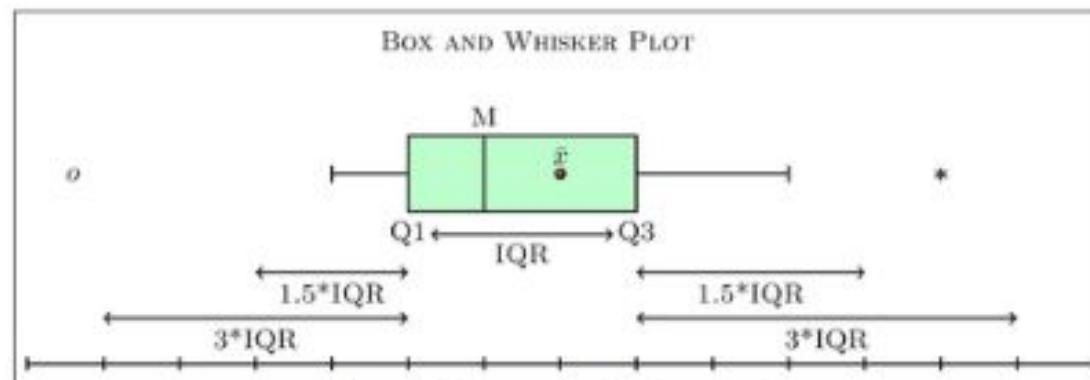


Figura 2. Ambiente para coleta de dados no apiário Emil.  
Fonte: <http://www.evit.no/wp-content/uploads/2016/05/emil-scale2.jpg>

# Preprocessing

- (i) Anomaly removal was performed by the Tukey Method;

$$[Q_1 - 1,5 \times (Q_3 - Q_1), Q_3 + 1,5 \times (Q_3 - Q_1)]$$



- (ii) Normalization was performed by the standard score (z-score);

$$Z = \frac{x - \mu}{\sigma}$$

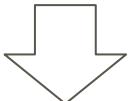
# Pattern recognition strategies

- (i) **clustering** (unsupervised approach)

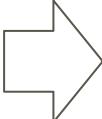
- Clustering Validation (CH index)
- Clustering (K-means, SSD)

$$CH(K) = \frac{tr(B_K)/(K - 1)}{tr(W_K)/(N - K)}$$

$$SSD = \sum_{\forall x \in V} |x - w|^2$$



- (ii) **analysis and interpretation of groups by an expert** (semi-supervised learning)



- (iii) **classification** (supervised approach) algorithms:
  - Naive Bayes
  - k-Nearest Neighbors(k-NN)
  - Random Forest

# Experiment Setup

- Removal of anomalies (cleaning) and normalization of data;
- Datasets division in 2 periods:
  - March/17 a August/17 – spring/summer;
  - September/17 a February/18 – fall/winter;
- K-means application and evaluation of the best prototypes by SSD ( $K_{max} = 24$ );
- Application of the Calinski-Harabasz (CH) index to set the most appropriate K ( $K_{opt}$ );

# Experiment Setup

---

**Algorithm 1:** Methodology for Application of K-means Clustering.

**Data:** Dataset (cleaned and normalized)

**Result:** Clusters

1. Calculate CH index for a set of  $K$  clusters.

$$K = \{2, \dots, k_{max}\}$$

**for each**  $k \in K$  **do**

- 1.1 Apply the k-means algorithm for  $N_r$  rounds.

- 1.2 Choose the prototypes of rounds with smaller  $SSD$ .

- 1.3 Calculate the value of the CH index for  $k$ .

**endfor**

2. Choose optimal value  $k_{opt}$  which optimizes the CH index.

3. Partition the data between  $k_{opt}$  groups using the criterion of Euclidean distance to the nearest prototype.

4. Report attribute descriptive statistics by grouping.

# Clustering Results

## Centroids

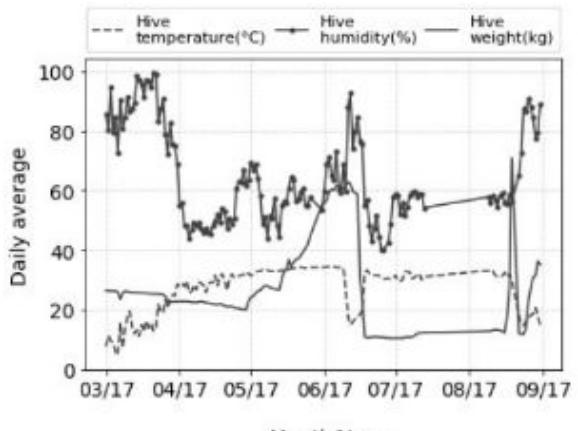
	Arnas						Emil					
	1° period			2° period			1° period			2° period		
	T(°C)	RH(%)	W(Kg)									
C0	14.3	89.2	25.3	10.3	78.1	31.0	26.5	55.0	12.4	13.5	79.9	15.6
C1	29.9	53.3	23.3	13.4	93.3	31.2	29.4	69.1	14.3	7.2	82.1	14.5
C2	30.8	54.0	11.1	7.7	96.0	29.5	10.9	75.8	14.0	7.0	80.6	16.5
C3	33.6	58.4	37.3							5.9	74.4	14.8
C4	16.4	85.9	24.6							7.9	78.4	15.2
C5										16.3	74.2	16.9

Table 1: Centroids of clusters

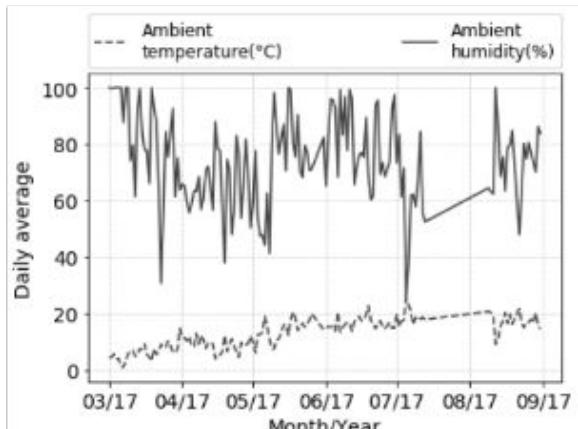
# Discussion

## Arnas 1º período

	°C	%	kg
14.3	89.2	25.3	
29.9	53.3	23.3	
30.8	54.0	11.1	
33.6	58.4	37.3	
16.4	85.9	24.6	

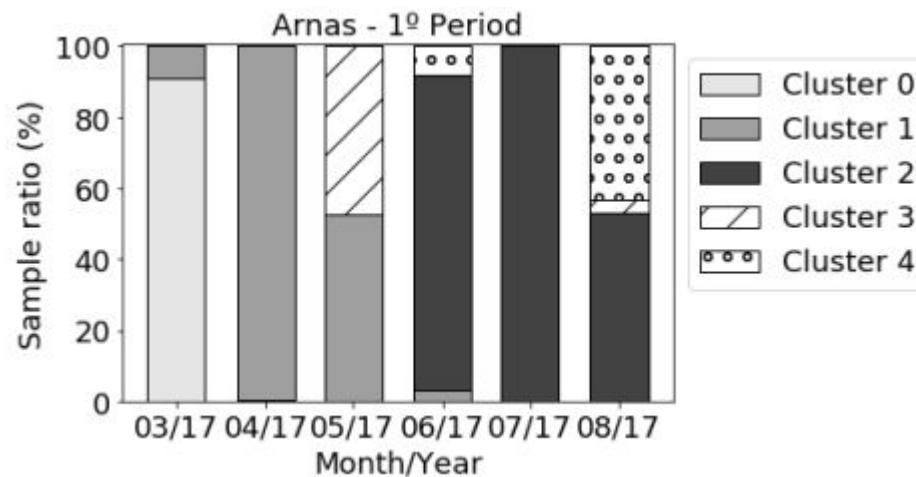


(a) Internal beehive characteristics.



(b) External beehive characteristics.

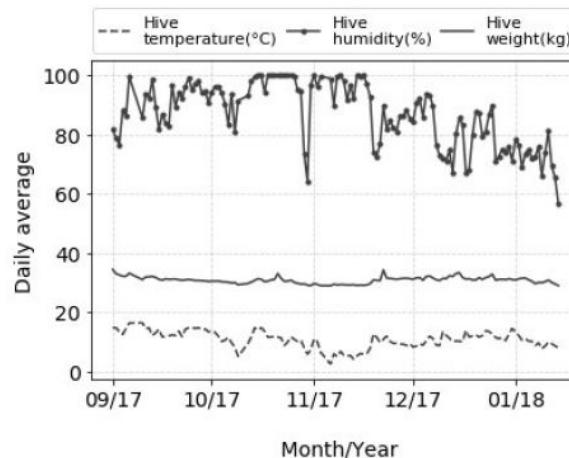
- C0 - Begin the brood rearing**  
High temperature and higher humidity
- C1 - Breeding and food collection**  
Low relative humidity
- C3 - Apex honey production and preparation for winter**  
Highest value of temperature and a higher value of HR
- C2 - Honey extraction**  
Low value of the mass
- C4 - Preparation to overwinter**  
Higher temperature and a lower humidity than cluster 0



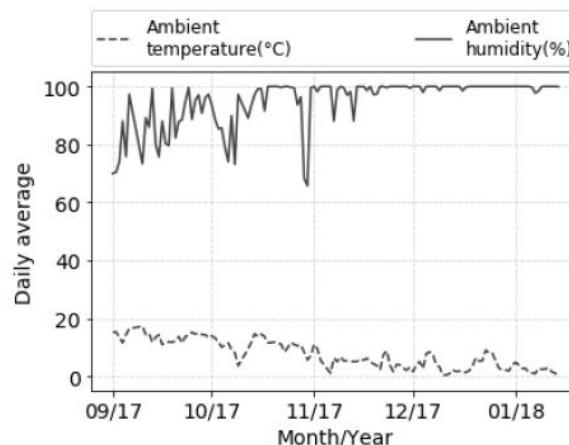
# Discussion

## Arnas 2º período

	°C	%	kg
10.3	78.1	31.0	
13.4	93.3	31.2	
7.7	96.0	29.5	

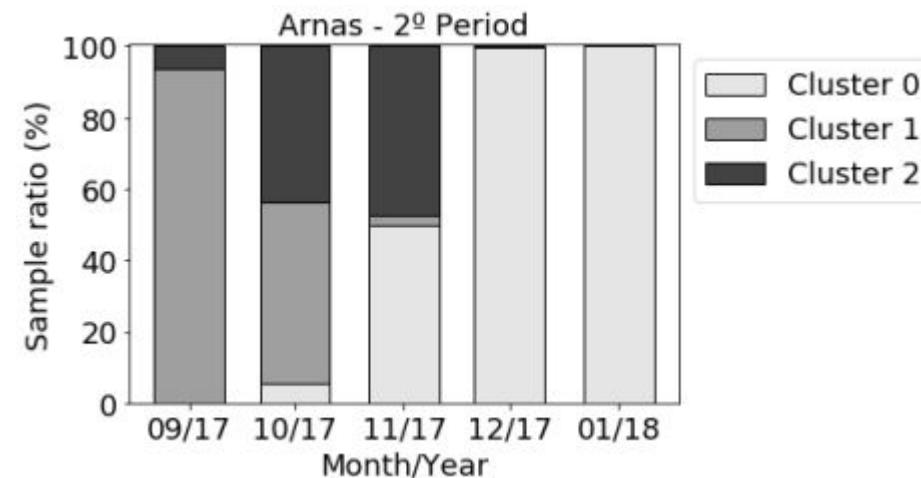


(a) Internal beehive characteristics.



(b) External beehive characteristics.

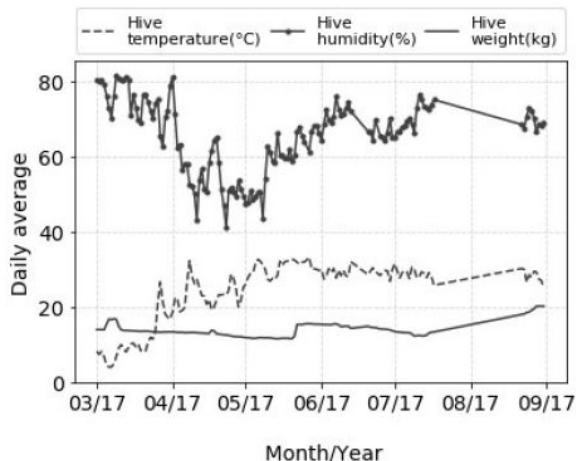
- **C0 - Thermoregulation (winter to spring)**  
Temperature inside the hive higher than the outside
- **C1 - Thermoregulation (autumn to winter)**  
Low relative humidity
- **C2 - Overwinter**  
Lowest temperature and highest relative humidity



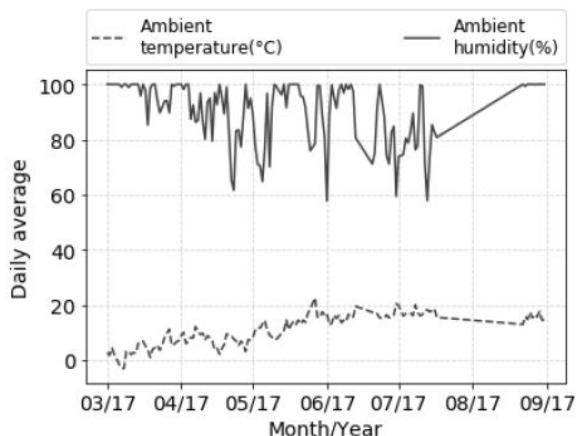
# Discussion

## Emil 1º período

	°C	%	kg
26.5	55.0	12.4	
29.4	69.1	14.3	
10.9	75.8	14.0	

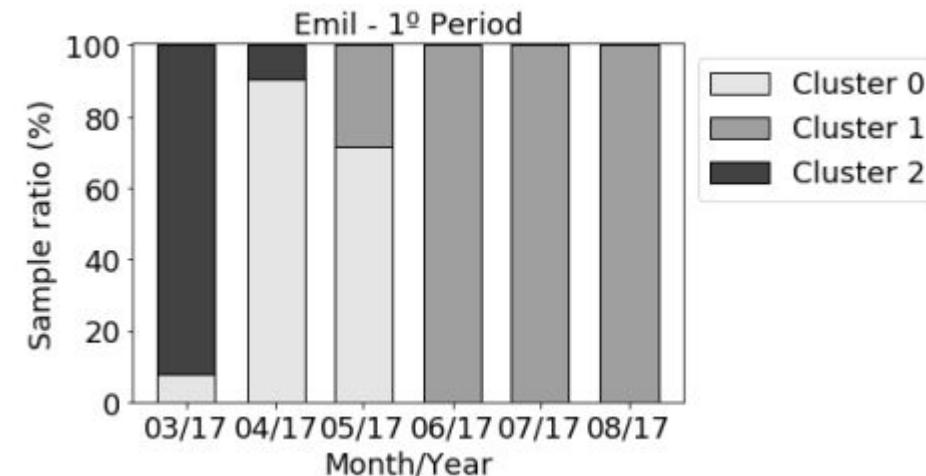


(a) Internal beehive characteristics.



(b) External beehive characteristics.

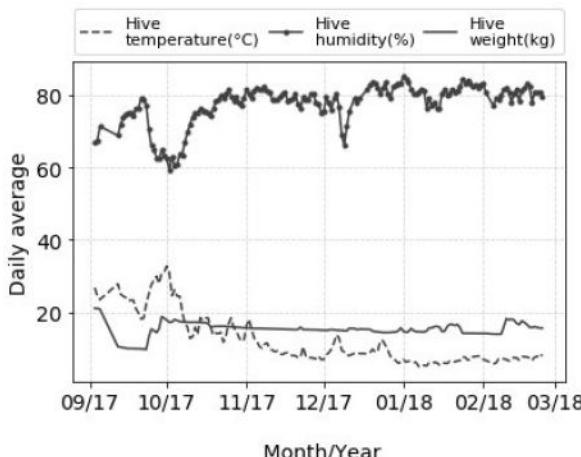
- **C2 - Begin the brood rearing (winter to spring)**  
Lowest temperature value and highest RH value
- **C0 - Intensification of breeding, collection and storage of food**  
Low relative humidity
- **C1 - The apex of honey production and the preparation for overwinter**  
Highest value of temperature and a higher value of HR



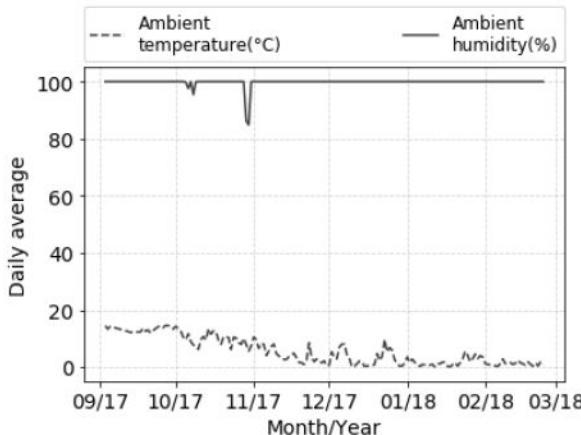
# Discussion

## Emil 2º período

	°C	%	kg
13.5	79.9	15.6	
7.2	82.1	14.5	
7.0	80.6	16.5	
5.9	74.4	14.8	
7.9	78.4	15.2	
16.3	74.2	16.9	

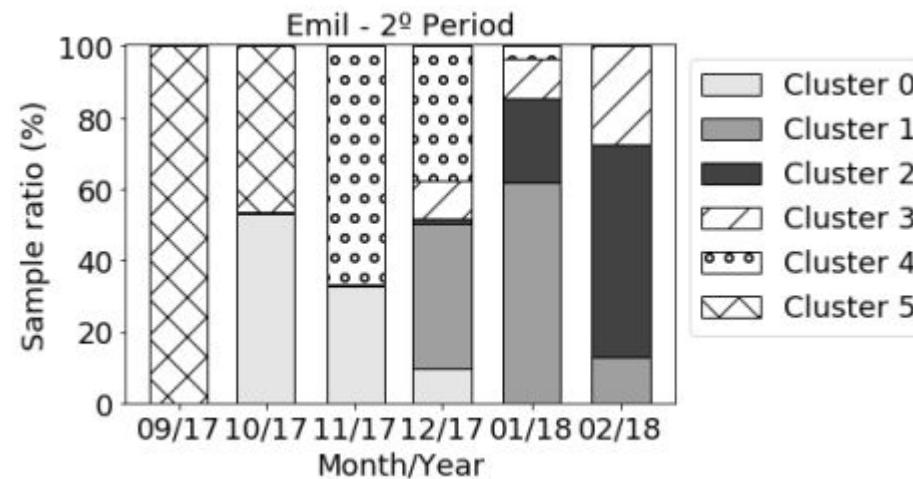


(a) Internal beehive characteristics.



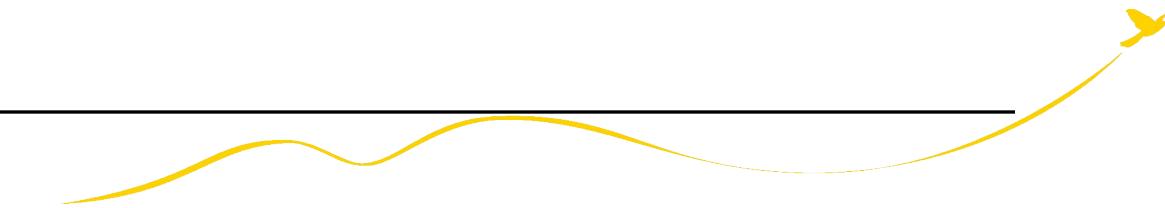
(b) External beehive characteristics.

- **C5 e C0 - Thermoregulation**  
Temperature is still not too low
- **C4 e C3 - Overwinter**  
Low Temperature
- **C2 e C1 - Overwinter**  
Increase in temperature



# Discussion

- Summary of identified states:
  - Beginning of brood rearing;
  - Brood rearing and food collection;
  - The apex of food storage;
  - Honey extraction;
  - Preparation to overwinter;
  - Overwinter.



# Discussion



The model generated by Random Forest was the most accurate, with an average accuracy of 98.77%.

Algorithm	Apiary				Average	
	Arnas		Emil			
	1st Period	2nd Period	1st Period	2nd Period		
N. Bayes	93.71 ± 0.31	88.41 ± 0.38	96.76 ± 0.19	92.04 ± 0.29	92,73%	
k-NN	98.96 ± 0.50	96.53 ± 1.18	99.63 ± 0.65	99.30 ± 0.56	98,60%	
R. Forest	99.13 ± 0.56	97.07 ± 1.08	99.67 ± 0.41	99.24 ± 0.86	98.77%	

Table 6: Success rates of classification algorithms.

# Conclusions



## Main aim

- To characterize the bee colony patterns *Apis mellifera* according to the changes of the seasons;

## Main contribution

- A Cluster-Classification Method for Accurate Mining of Seasonal Honey Bees Colonies Patterns;

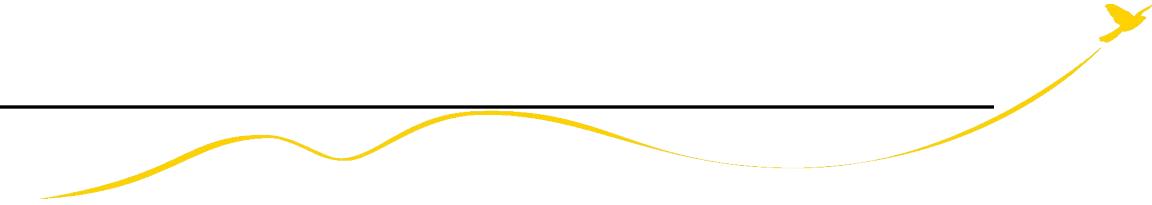
## Results

- The patterns found were recognized as consistent;
- The apparently stronger beehive performs more effectively thermal control during the winter;
- A high-accuracy classification model for new samples;

# Questions?

19/20





# Thank you!

[antoniorafaelbraga@gmail.com](mailto:antoniorafaelbraga@gmail.com)