

Análisis de componentes principales utilizando python para identificar clúster asociados a muestras de cacao seco sano e infectado con monilia en Norte de Santander.

Principal component analysis using python to identify clusters associated with samples of healthy monilia-infected dry cocoa in Norte de Santander.

Alexander Flórez-Martínez¹, Jesus Omar Vargas-Flórez², Harold Esneider Perez-Waltero³,
Lucas Fernando Quintana-Fuentes⁴
^{1,2,3,4}Universidad Nacional Abierta y a Distancia, Colombia

Recibido: 13 de enero de 2020.

Aprobado: 13 de abril de 2020.

Resumen— El resultado del proyecto de investigación está asociado a data proveniente de la implementación de un sistema electrónico multisensorial o común mente denominado olfato electrónico. Mediante el uso de un sistema de adquisición de datos y software LabView se almacena la data de la concentración de volátiles asociados a muestras de cacao sano e infectado con monilia, aplicado en la fase o etapa de secado. Una vez adquirido los datos se procede a implementar en el software Python el pre-procesamiento y procesamiento de data, permitiendo al usuario por medio de un gráfico identificar el clúster asociados a cada clase, cacao sano o con monilia. Como método para aprendizaje de automático no supervisado, se implementa análisis de componentes principales PCA para el respectivo procesamiento. Los resultados obtenidos varían de acuerdo al método de preprocesado de datos. Para el desarrollo se implementó un escalador robusto y preprocesado euclidiano, el cual presenta mejores resultados de agrupamiento de muestras por clase.

Palabras Claves: Olfato electrónico, Procesamiento de datos, Aprendizaje no supervisado, Software Python, secado de cacao.

Abstract— The result of the research project is associated with data from the implementation of a multisensory electronic system or commonly called electronic smell. Through the use of a data acquisition system and LabView software, the volatile concentration data associated with samples of healthy cocoa and infected with monilia, applied in the drying phase or stage. Once the data is acquired, the Python software used for preprocessing and processing data. Allowing the user to identify the cluster associated with each class, healthy cocoa or with monilia. As a method for learning unsupervised automatic, PCA principal component analysis is implemented for the respective processing. The results obtained vary according to the method of data preprocessing, a robust climber and preprocessed Euclidean was implemented, which presents better results of grouping samples by class.

Keywords: Electronic smell, Data processing, Unsupervised learning, Python software, cocoa drying.

*Autor para correspondencia.

Correo electrónico: Alexander.florez@unad.edu.co (Alexander Flórez Martínez).

La revisión por pares es responsabilidad de la Universidad de Santander.

Este es un artículo bajo la licencia CC BY-ND (<https://creativecommons.org/licenses/by-nd/4.0/>).

Forma de citar: A. Flórez-Martínez, J. O. Vargas Flórez, H. E. Perez-Waltero y L. F. Quintana-Fuentes, "Análisis de componentes principales utilizando python para identificar clúster asociados a muestras de cacao seco sano e infectado con monilia en Norte de Santander", Aibi revista de investigación, administración e ingeniería, vol. 8, no. 2, pp. 16-22, 2020.

I. INTRODUCCIÓN

En los últimos años se ha planteado el uso de instrumentos analíticos que intentan emular el proceso de identificación de aromas del que dispone el cuerpo humano mediante la construcción de sistemas de olfato electrónico, más conocidos como narices electrónicas, para la extracción de los rasgos característicos de los alimentos, porque ofrecen una alternativa rápida y no destructiva para detectar el aroma, y por ende, pueden ser usadas en diferentes campos con ventajas técnicas y económicas frente a otros procedimientos como la cromatografía de gases o el análisis sensorial [1].

En la actualidad las narices electrónicas no solo se usan para clasificaciones de aroma, otra alternativa es la detección de olores. De ahí que su importancia y utilización se haya extendido a aplicaciones en la agroindustria, medio ambiente, la seguridad, y la medicina [2].

Un dispositivo de este tipo lo componen principalmente 4 etapas con diferentes funciones: 1) La primera realiza la adecuación de la mezcla gaseosa y el muestreo. 2) El conjunto de sensores de gases hace la detección de los compuestos volátiles. 3) La electrónica de control se dedica a la gestión del conjunto de sensores y adecuación de la señal. 4) Finalmente, en una computadora se extraen los rasgos característicos o "huellas" de cada aroma, y se aplican algoritmos de clasificación de patrones [3].

El tratamiento de los datos constituye uno de los pilares fundamentales de un sistema de nariz electrónica capaz de detectar, cuantificar e identificar compuestos volátiles. Un buen diseño de un sistema de reconocimiento de patrones para una nariz electrónica requiere una optimización de los distintos procedimientos involucrados en el procesamiento de los datos: preprocesado de la señal, extracción y selección de características, clasificación y validación [4].

Algunos de los sentidos de los seres vivos se han implementado con bastante éxito de manera artificial mediante sistemas electrónicos, de esta forma se han descrito aplicaciones de visión artificial reconocimiento de caracteres, reconocimiento de voz. El sentido del olfato ha sido uno de los que menos desarrollo artificial ha tenido en este aspecto [5].

El desarrollo de la investigación aborda el análisis de compuestos volátiles en el proceso de secado de cacao, que aporta significativamente en los precursores de fino aroma del grano.

Basado en un análisis organoléptico más confiable, es decir, una nariz electrónica (EN) para identificar defectos en las muestras como lo hacen los catadores, pero basados en patrones de capacitación para que el análisis y los resultados no se vean afectados por factores externos [6].

Si el producto no cumple con el aroma requerido, conlleva a que el grano se considere de baja calidad y se pague a menor precio o se considere no apto para comercialización. En algunos casos el veredicto de aroma se somete a la subjetividad humana que se relaciona con la percepción que emite el ser humano, antes de hacer la cata se debe evitar: fumar, tomar alcohol y café, estar trasnochado o con gripa, ya que este tipo de situación hace que se bloquee el sistema sensorial [7] y no se pueda dar un veredicto confiable del control de calidad del producto. Existen otros métodos para determinar la calidad del grano como la cromatografía de gases y espectrometría de masas, pero se requiere pagar un precio adicional por cada muestra y este sometido a un tiempo de respuesta debido a que el grano a analizar y enviarlo a un laboratorio especializado y pocas instituciones prestan este servicio. Los resultados de la investigación muestran como un sistema de olfato electrónico por medio de algoritmos basados en inteligencia artificial, puede llegar a discriminar olores entre muestras de cacao seco deseado y muestras de cacao seco infectado con monilia. Otra de las ventajas del sistema de olfato electrónico son la rapidez de medida, confiabilidad y fácil de usar. Actualmente continuamos trabajando en

la optimización del sistema, entrenamiento, en camino para llegar a una solución adecuada para la industria.

En la sesión II se define el marco teórico, relacionando diagrama de bloques de una nariz electrónica, características de los sensores de gases, procesamiento de datos por análisis de componentes principales PCA. En la sesión III se explica la metodología o procedimientos se explica el protocolo implementado para la fase secado de cacao, selección, preparación de muestras y adquisición de datos. En la sesión IV se socializa resultados, análisis e interacción de clúster asociados a muestras de cacao seco sano e infectado con monilia. En la sesión V se explica propuesta de solución y mejoras. Finalmente, en la sesión VI se describe las conclusiones.

II. MARCO TEÓRICO

El diseño de instrumentos basados en los mismos principios que los sentidos asociados a la percepción humana del entorno ha sido objeto de gran interés en investigación en diversas áreas de la ingeniería durante décadas. Comparativamente, sin embargo, solo algunos de estos sentidos, como la visión artificial, han alcanzado un nivel de madurez tal que permite hoy su uso habitual en la industria. El reto de replicar artificialmente el sentido olfativo biológico ha dado lugar a sistemas electrónicos compactos conocidos como narices electrónicas o e-noses, en inglés. Estos sistemas requieren, por un lado, el uso de sensores diseñados específicamente para detectar olores, y por otro, la aplicación de técnicas de software de inteligencia artificial para el procesamiento de la información y la clasificación de los datos en tiempo real. Algunos trabajos implementan el sistema de olfacción artificial basado en servicios web, que posibilita la conexión remota en línea de narices electrónicas para el envío de sus datos adquiridos, y la recepción de los resultados de clasificación de aromas de interés, obtenidos mediante técnicas de reconocimiento de patrones sobre dichos datos, con mínimo retardo [8].

En los sistemas multisensoriales, específicamente para el funcionamiento de las narices electrónicas actuales, el computador es el más utilizado para realizar el procesamiento y la clasificación de los datos provenientes de una matriz de sensores de gases químicos [9].

Los sistemas de nariz electrónica se diseñaron para ser usados con numerosos productos y aplicaciones, como alimentos, empaques, cosméticos, control de calidad de materias primas y productos elaborados; monitoreo de procesos, frescura y madurez, investigaciones de estimación del tiempo de vida útil, valoraciones de autenticidad de productos, clasificación de esencias y perfumes, detección de microorganismos patógenos y estudios medioambientales [10].

El desarrollo y la optimización de métodos de control en cada etapa del proceso de producción, desde las materias primas hasta los productos finales, se está volviendo cada vez más importante en la tecnología de procesamiento de alimentos. Existe un creciente interés en los temas de evaluación de la inocuidad y la calidad de los alimentos, que se han convertido en uno de las principales prioridades en el análisis de alimentos [11], se prevé la introducción de narices electrónicas en el área de alimentos para control de calidad, monitoreo de procesos, evaluación de frescura, investigación de vida útil y evaluación de autenticidad [12]. El campo de tecnología de sensores avanza durante los últimos veinte años, en gran parte debido al descubrimiento de numerosas aplicaciones derivadas de las investigaciones en diversos campos de las ciencias aplicadas [13].

Los sistemas de olfato electrónico en su estructura cuentan con algún tipo de mecanismo de limpieza de la cámara de sensores de forma que las medidas sucesivas se hagan partiendo de las mismas condiciones iniciales y se garantice la repetibilidad de los resultados. Se recomienda utilizar otra cámara o compartimento hermético para alojar la muestra que se desea analizar, siempre que las condiciones

Análisis de componentes principales utilizando python para identificar clúster asociados a muestras de cacao seco sano e infectado con monilia en Norte de Santander del medio y del experimento lo permitan [14], de esta forma los sistemas de olfato electrónico son equipos automatizados no destructivos que permiten caracterizar el olor y el sabor en los alimentos [15].

En el diagrama de bloques de la Figura 1 se ilustra diagrama de bloques de un sistema de olfato o nariz electrónica [5].

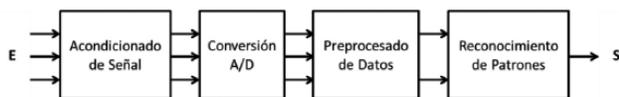


Figura 1: Diagrama de bloques de una nariz electrónica. Fuente: [8].

El primer bloque de la Figura 1, acondicionamiento de señal, para los sensores de gases normalmente se requiere una tarjeta de acondicionamiento que traduzca la medición de compuestos volátiles en una señal de salida equivalente en voltaje.

La primera etapa según el diagrama de bloques de la nariz electrónica relaciona el acondicionamiento de señal, en el caso de los sensores de gases algunos incluyen una tarjeta de acondicionamiento que permite realizar medición de compuestos volátiles y transformarlos a una señal de salida equivalente en voltaje.

Sensores de gas.

Los detectores de gas miden e indican la concentración de ciertos gases en el aire a través de diferentes tecnologías y estos detectores pueden clasificarse según el tipo de gas que detectan (inflamable o tóxico). Antiguamente se diseñaban para detectar un único gas. Ahora existen múltiples dispositivos multifuncionales capaces de detectar varios gases a la vez.

Actualmente, en función del tipo de gas, podemos resumir las siguientes tecnologías:

Para detección y medición de gases tóxicos:

Sensores electroquímicos: son los más usados para la detección de gases tóxicos como monóxido de carbono, cloro y óxidos de nitrógeno. Funcionan por medio de señales de electrodos cuando se detecta el gas.

Semiconductores de óxido de metal (o MOS): también se usan para detectar gases tóxicos (normalmente monóxido de carbono, aunque es capaz de detectar también sulfuro de hidrógeno, H₂S). Trabajan a través de una película sensible al gas, está compuesta de óxido de estaño o tungsteno. Estos sensores suelen considerarse eficientes por su capacidad para trabajar en rangos de baja humedad, además de ser capaces de detectar también gases combustibles.

Sensores catalíticos: representan un gran número de los dispositivos detectores de gas que se fabrican hoy en día. Se usan para detectar gases tales como hidrocarburos, funcionando a través de la oxidación catalítica. A pesar de que pueden ser calibrados para medir varios tipos de gases, el metano es el gas más común que suelen detectar. La salida de un sensor catalítico se mide mediante un dispositivo llamado puente de Wheatstone (usado comúnmente en circuitos eléctricos), que se usa para medir la resistencia eléctrica desconocida, que puede ser analizada para proporcionar la información adecuada en relación con los gases presentes.

- **Sensores de infrarrojos o detectores de IR:** funcionan a través de un sistema de transmisores y receptores para detectar gases combustibles (concretamente vapores de hidrocarburos). Si el gas está presente en la trayectoria óptica, se interfiere el poder de transmisión de luz entre el transmisor y el receptor. La alteración del estado de la luz determina si hay gas y además qué tipo de gas está presente. Estos sensores en lugar de depender de un puente de Wheatstone para medir

la diferencia de la resistencia eléctrica, emplean una lámpara de fuente de infrarrojos y un filtro de longitud de onda [16].

Algunos sensores de tipo electrolítico comerciales son: el MQ-2, MQ-3, MQ-7 y MQ-135.

Sensor MQ-2: es un sensor ideal para detectar la presencia de una peligrosa fuga de gas licuado de petróleo. Se puede integrar en diferentes aplicaciones de automatización que estén asociadas al internet de las cosas, por ejemplo, en el hogar, en una estación de servicio, ambiente de tanque de almacenamiento. El sensor tiene una excelente sensibilidad con un rápido tiempo de respuesta. La conductividad del sensor es mayor a medida que aumenta la concentración de gas [17].

Sensor MQ-3: tiene una alta sensibilidad al ser expuesto ante compuestos volátiles como: alcohol, humo. El sensor podría usarse para detectar alcohol con diferentes concentraciones Es de bajo costo y adecuado para diferentes aplicaciones [18].

Sensor MQ-7: El sensor de gas es un semiconductor de monóxido de carbono, es adecuado para detectar concentraciones de CO en el aire. Puede detectar concentraciones de COgas que oscilan entre 20 y 2000 ppm. Este sensor tiene una alta sensibilidad y un tiempo de respuesta rápido [19]

Sensor MQ-135 El sensor de gas MQ135 tiene alta sensibilidad al amoníaco, sulfuro y vapor Benz E, también sensible al humo y otros gases nocivos Cada sensor [20].

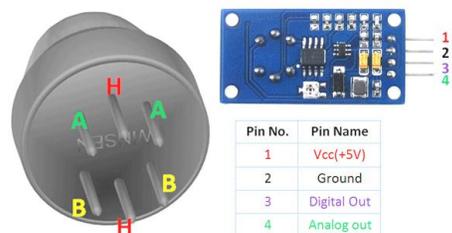


Figura 2: Partes de un sensor MQ con tarjeta de acondicionamiento integrada. Fuente: Módulo del sensor MQ [21].

La tabla 1, resumen las características de los sensores MQ-2, MQ-3, MQ-7, MQ-135.

Tabla 1: Características de los sensores de gases MQ.

Sensor	Otros compuestos químicos detectados
MQ-2	H ₂ , LPG, CH ₄ , CO
MQ-3	Alcohol, Benceno, CH ₄ , Hexano
MQ-7	Monóxido de Carbono, H ₂
MQ-135	NH ₃ , NOx, alcohol, benceno, humo, CO ₂

Fuente: sensores de gases MQ [22].

Python para procesamiento de datos

Python es un lenguaje de programación de alto nivel, interpretado y multipropósito. Python puede utilizarse en varios sistemas operativos como Windows, Mac OS y Linux. Este lenguaje es considerado potente, flexible y con una sintaxis clara y concisa [23].

Scikit-learn es un módulo de Python que integra una amplia gama de algoritmos de aprendizaje automático de última generación para problemas supervisados y no supervisados de mediana escala. Este paquete se centra en llevar el aprendizaje automático a los no especialistas que utilizan un lenguaje de alto nivel de propósito general [24]

Hoy nos encontramos en los prolegómenos de la cuarta revolución industrial, conocida como Industria 4.0. Bajo este término, se busca introducir en la industria nuevas tecnologías como pueden ser la

inteligencia artificial, el análisis de datos, la realidad aumentada, la realidad virtual, el Big Data, el Cloud o el Internet of Things [25].

En el desarrollo del proyecto se utiliza el entorno de desarrollo Spyder, el cual integra librerías para programación en Python. Con la data adquirida resultado de la toma de muestras de los compuestos volátiles de cacao seco, se procede a realizar en Python, extracción de parámetros, posterior inicia el preprocesamiento y procesado de datos. En la etapa de preprocesado, se utilizaron las funciones `robustScaler` y `euclidiano`. Finalmente se grafican los cluster mediante procesado por análisis de componentes principales y se logran discriminar las muestras asociadas a cacao seco sano e infectado con monilia.

Método de procesamiento de datos mediante análisis de componentes principales PCA

PCA es un método que asume colinealidad (característica de un conjunto de puntos situados sobre la misma recta), entre las variables que intervienen. En otras palabras, se trata de un algoritmo lineal que puede funcionar incorrectamente en procesos altamente no lineales, como puede ser las iteraciones químicas entre sensores y compuestos volátiles.

PCA funciona sorprendentemente bien en muchas aplicaciones con sistemas de olfato electrónico [1].

PCA es un algoritmo de aprendizaje no supervisado ya que las instrucciones de estos componentes se calculan exclusivamente a partir del conjunto de características explicativas sin ninguna referencia a las variables de respuesta. El número de combinaciones de características es igual al número de dimensiones del conjunto de datos y, en general, establece el número máximo de PCA que se pueden construir.

En la figura 3, cada punto azul corresponde a una observación, y cada componente principal reduce las tres dimensiones a dos. El algoritmo encuentra un par de vectores ortogonales (flechas rojas) que definen un espacio de menor dimensión (plano gris) para capturar la mayor variación posible del conjunto de datos original [26].

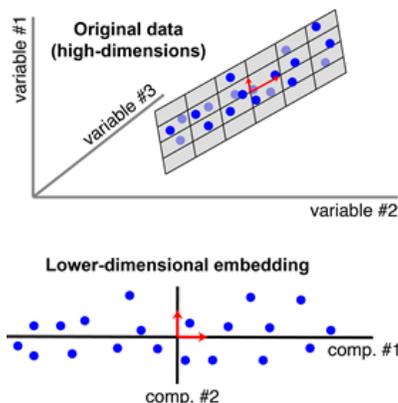


Figura 3: Análisis de componentes principales.
Fuente: Vectores ortogonales [26].

Sistema de adquisición de datos basado en hardware Arduino

Arduino es una plataforma de desarrollo de microcontrolador unida a un lenguaje de programación intuitivo basado en C/C++. Equipando Arduino con sensores, actuadores es posible convertir la plataforma en un “cerebro” programable para casi cualquier sistema de control [16].

La plataforma de desarrollo para la Adquisición de Datos fue Arduino Mega, dispositivo encargado de recibir las señales analógicas provenientes de 9 sensores de gases MQ, con las siguientes referencias:

MQ4, MQ135, MQ2, MQ8, MQ5, MQ6, MQ3, MQ7, MQ9 alojados en una cámara de medida, para operar el sistema de adquisición se realizó una interfaz gráfica con el software LabView quien es el encargado de guardar la data en Excel, posteriormente se utiliza Python, para el procesamiento de datos con análisis de componentes principales PCA.

La matriz de sensores que conforman la cámara de medida se describe en la tabla 2.

Tabla 2. Sensores MQ que conforman la cámara de medida.

MQ-2	Sensor adecuado para detectar GLP, propano, metano, alcohol, hidrogeno, humo.
MQ-3	Muy sensible al alcohol y de menor sensibilidad a la bencina, también es sensible a gases como GLP, Hexano, CO, CH4 pero con sensibilidad muy baja, la cual se puede despreciar si hay poca concentración.
MQ-4	Gas Metano, puede detectar concentraciones desde las 300 hasta las 10000 ppm.
MQ-5	Detección de GLP, propano, butano
MQ-6	Es adecuado para detectar la presencia de Gas LP, compuesto principalmente por Propano y Butano y Gas Natural (Metano) en el aire. El sensor puede detectar concentraciones de gas entre 200 y 1000 ppm. Otros compuestos: Monóxido de carbono y alcohol.
MQ-7	Sensor de alta sensibilidad al monóxido de carbono (CO), pero también es sensible al H2.
MQ-8	Adecuado para detectar concentraciones de hidrógeno en el aire. Puede detectar concentraciones de gas de hidrógeno en cualquier lugar de 100-10000 ppm.
MQ-9	Sensor de Monóxido de Carbono (CO) y de gas inflamable.
MQ-135	Se utilizan en equipos de control de calidad del aire, son adecuados para la detección de NH3, NOx, alcohol, benceno, humo, CO2, etc.

Fuente: Sensores MQ [27].

III. METODOLOGÍA O PROCEDIMIENTOS

El proceso inicia en la recolección de la mazorca de cacao del material ICS-95 de origen trinitario, se caracteriza por ser de fino aroma, se procede a realizar el protocolo sugerido por FEDECACAO en el proceso de fermentación, posterior el proceso de secado el cual se sigue las siguientes indicaciones:

Protocolo de secado del cacao

Objetivo reducir la humedad al grano de cacao. Después de fermentado queda con humedad del 55%, por tanto, debe bajar hasta obtener del 7% al 8% de humedad a final del proceso de secado.

- Primer día: tres (3) horas de sol, luego la masa se recoge.
 - Segundo día: cuatro (4) horas de sol, luego se recoge.
 - Tercer, cuarto y quinto día, todo el día recibiendo radiación solar
- Se debe realizar remoción de la masa de cacao cada (2) horas, al menos en 4 y 5 remociones diarias.

Durante el proceso de recolección se encontraron algunas mazorcas infectadas con monilia, se procedió a realizar la preparación de la muestra hasta llegar a la etapa de secado con el objetivo de clasificar los compuestos volátiles emitidos por las almendras de cacao infectadas. La figura 4, ilustra muestras de cacao en el proceso de secado al sol.



Figura 4: Fase de secado de cacao ICS-95, CCN-51, Infectado con monilia. Ubicación UNAD Cúcuta. Fuente: Elaboración propia.

Aplicativo de software NOSE V1 para adquisición de datos

Para la adquisición de la huella olfativa proveniente de las muestras de cacao, se implementó (1) un sistema multisensorial basado en 9 sensores de gases del fabricante MQ. (2) Una tarjeta Arduino Mega para adquirir las señales de los sensores y enviar la información por protocolo USB al equipo de computo, donde se almacena la data relacionada a las diferentes muestras de cacao seco. El software para el desarrollo de la interfaz gráfica de usuario es LabView versión 2018, Finalmente la data se almacena en archivo tipo Excel, para su posterior procesamiento de datos con el uso del software Python. La figura 5, ilustra diseño de la interfaz gráfica de usuario para la operación y adquisición de datos del sistema de olfato electrónico.

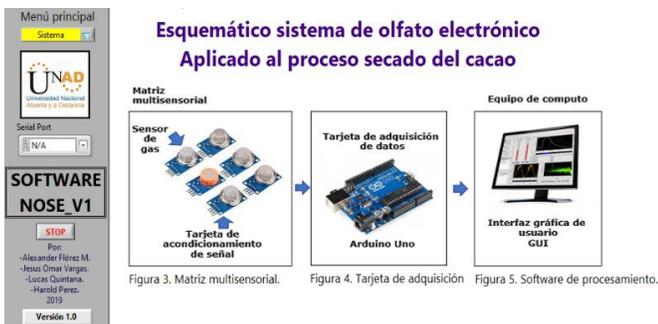


Figura 3. Matriz multisensorial. Figura 4. Tarjeta de adquisición. Figura 5. Software de procesamiento.

Figura 5: Sistema de olfato electrónico aplicado al proceso de secado del cacao. Fuente: Elaboración propia.

Se implementa el uso de la tarjeta de adquisición hardware arduino 2560, debido al número de canales y terminales de entrada/salida. La comunicación se establece mediante puerto USB a PC, el sistema es flexible en programación y es compatible con diversos tipos de software como LabView, Python. Para el desarrollo del sistema de olfato electrónico la comunicación se realiza con el software LabView, y uso de librerías MakerHub [28].

Las muestras se alistan en 10 viales herméticos, cada una con un peso 20 gramos del grano seco, 5 viales contienen granos de cacao sano ICS-95 y 5 viales continen granos de cacao infectado con monilia. El volátil se extrae por medio de una jeringa con un volumen de 20ml para proceder a inyectar en la cámara de concentración para el respectivo análisis de la huella olfativa.

IV. RESULTADOS, ANÁLISIS E INTERPRETACIÓN

Python es un lenguaje de programación que cumple con lo planteado y se viene perfilando como una opción recomendada para el desarrollo de software libre [29].

La data adquirida es almacenada en un formato Excel, y luego desde Python se lee y se prepara la data para ser procesada. (1) se

realiza la extracción de parámetros. (2) se utilizan dos tipos de pre-procesado robustScaler y euclidiano. (3) Se realiza el procesado de datos mediante análisis de componentes principales PCA, utilizando los dos tipos de pre-procesado enunciados.

Preprocesado de datos RobustScaler

Como método para aprendizaje de automático, se implementa análisis de componentes principales PCA para el respectivo procesamiento. Los resultados obtenidos varían de acuerdo al método de preprocesado de datos, para el desarrollo se implementó un escalador robusto.

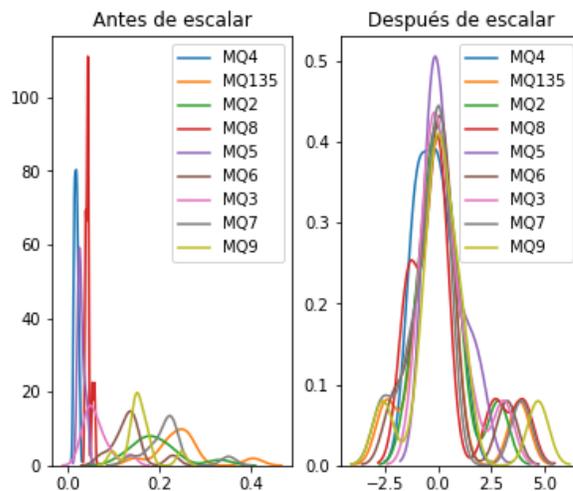


Figura 6: Preprocesado de datos RobustScaler con Python. Fuente: Elaboración propia.

Procesado de datos PCA (Análisis de componentes principales) en Python

La figura 7, ilustra clústeres asociados al proceso de secado de cacao mediante el uso de análisis de componentes principales preprocesado robusto $PC1 = 0.863$.

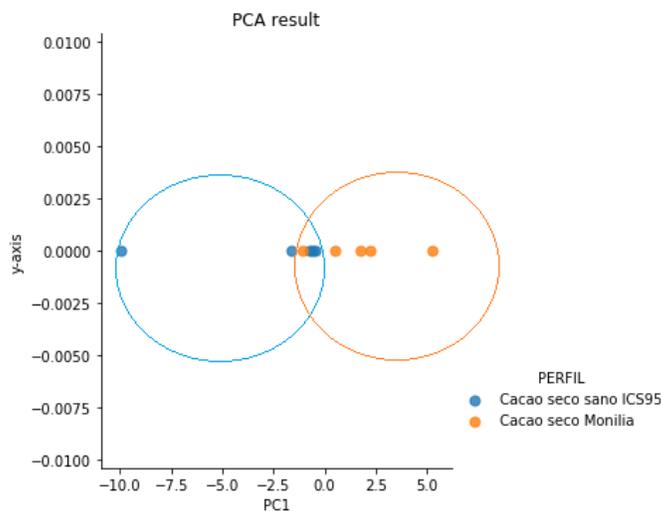


Figura 7: Clústeres asociados al proceso de cacao seco mediante análisis de componentes principales preprocesado escalador robusto con Python. Fuente: Elaboración propia.

0.863
0.077
Preprocesado de datos euclidiano

Como método para aprendizaje de automático, se implementa análisis de componentes principales PCA para el respectivo procesamiento. Los resultados obtenidos varían de acuerdo al método

de preprocesado de datos, para el desarrollo se implementó preprocesado euclidiano.

Euclídea o L2:

$$d_2(x, y) = \sqrt{\sum_{j=1}^J (x_j - y_j)^2} \quad (1)$$

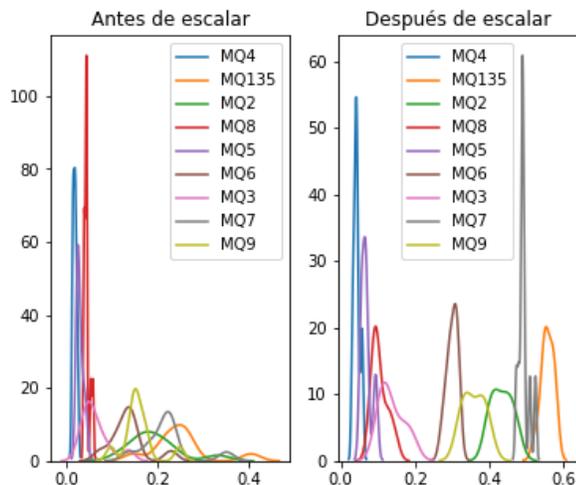


Figura 8: Preprocesado de datos euclidiano con Python. Fuente: elaboración propia.

Agrupamiento o clustering

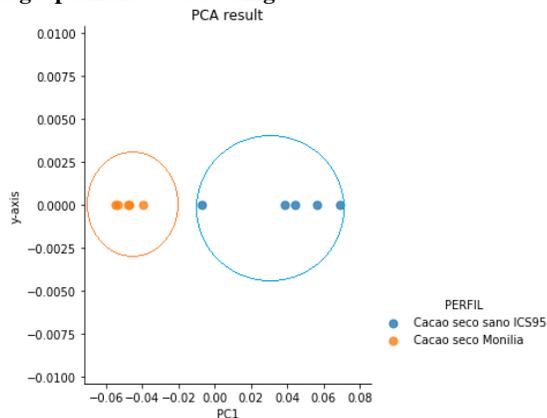


Figura 9: Clústeres asociados al proceso de cacao seco sano e infectado con monilia, en una componente mediante PCA pre-procesado euclidiano PC1=0.66.

Fuente: elaboración propia.

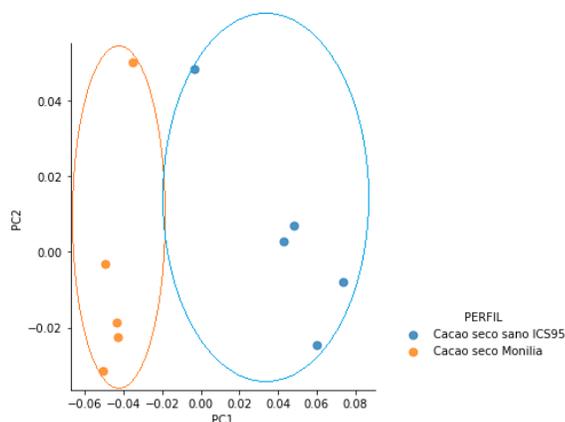


Figura 10: Clústeres asociados al proceso de cacao seco sano e infectado con monilia, en dos componentes mediante PCA pre-procesado euclidiano PC1=0.66, PC2=0.2.

Fuente: elaboración propia.

V. PROPUESTA DE SOLUCIÓN O MEJORAS

Se implementó un sistema inteligente, el cual es un sistema electrónico multisensorial denominado olfato electrónico que discrimina entre cacao sano e infectado con monilia.

El hardware utilizado fue Arduino, sensores MQ y los algoritmos se implementaron en Phytion. Como método para aprendizaje automático no supervisado, se utilizó análisis de componentes principales PCA para el respectivo procesamiento.

Los resultados obtenidos varían de acuerdo al método de preprocesado de datos. En el caso del escalador robusto se obtiene un porcentaje de varianza del 86%. En la primera componente principal se observa solapamiento de muestras entre las clases. Se procede a implementar el método de preprocesado euclidiano logrando el 66% de varianza en la primera componente principal y el 21% en la segunda componente, para un total de 87%. No se evidencia solapamiento entre las muestras de las clases, siendo el método más adecuado de clasificación.

VI. RECOMENDACIONES

La investigación se puede articular implementando técnicas de inteligencia artificial como redes neuronales, máquinas de soporte vectorial entre otros, que permitan entrenar el sistema de olfato electrónico para la clasificación de compuestos volátiles.

VII. REFERENCIAS

- [1] M. C. Quicazán y C. M. Zuluaga, «La nariz electrónica, una novedosa herramienta para el control de procesos y calidad en la industria agroalimentaria.», vol. 18, n.o 2, p. 9, 2011.
- [2] I. Moreno, R. Caballero, R. Galán, F. Matía, y A. Jiménez, «La Nariz Electrónica: Estado del Arte», Rev. Iberoam. Automática E Informática Ind. RIAI, vol. 6, n.o 3, pp. 76-91, 2009, doi: 10.1016/S1697-7912(09)70267-5.
- [3] A. C. Durán, «SISTEMA DE OLFATO ELECTRONICO PARA LA DETECCION DE COMPUESTOS VOLATILES», vol. 2, n.o 12, 2008.
- [4] J. Lozano et al., «IDENTIFICACIÓN DE COMPUESTOS AROMÁTICOS DEL VINO MEDIANTE NARIZ ELECTRÓNICA», p. 5.
- [5] J. Lozano, J. P. Santos, M. Aleixandre, y M. C. Horrillo, «IDENTIFICACIÓN DE AROMAS EN VINOS BLANCOS Y TINTOS MEDIANTE SENSORES DE ESTADO SÓLIDO», p. 5.
- [6] J. Rodríguez, C. Durán, y A. Reyes, «Electronic Nose for Quality Control of Colombian Coffee through the Detection of Defects in “Cup Tests”», Sensors, vol. 10, n.o 1, pp. 36-46, dic. 2009, doi: 10.3390/s100100036.
- [7] M. Martínez, «Catar cacao es cuestión de actitud pero también de aptitud», 2013. <http://www.fedecacao.com.co/portal/index.php/es/2015-04-23-20-00-33/126-catar-cacao-es-cuestion-de-actitud-pero-tambien-de-aptitud> (accedido may 29, 2020).
- [8] J.-Á. Fernández-Muñoz, J.-L. Herrero-Agustín, J. Lozano-Rogado, J.-P. Santos-Blanco, y J.-I. Suárez-Marcelo, «Un sistema de clasificación de olores en línea para narices electrónicas. (Spanish)», Online Odor Classif. Syst. Electron. Noses Engl., vol. 92, n.o 3, pp. 354-360, 5/1/2017 2017, doi: 10.6036/8075.
- [9] O. E. G. Guerrero, C. M. D. Acevedo, J. A. A. Gallardo, y J. E. O. Sandoval, «IMPLEMENTATION OF A NEURONAL MODEL IN A HARDWARE DEVICE (FPGA) FOR THE

CLASSIFICATION OF CHEMICAL COMPOUNDS IN A MULTISENSORY (E-NOSE)», p. 7, 2014.

- [10] Quicazán, «Electronic nose a novel tool for quality and process control in the food industry», vol. 18, p. 9, 2011.
- [11] W. Wojnowski, T. Majchrzak, T. Dymerski, J. Gębicki, y J. Namieśnik, «Portable Electronic Nose Based on Electrochemical Sensors for Food Quality Assessment», *Sensors*, vol. 17, n.o 12, p. 2715, nov. 2017, doi: 10.3390/s17122715.
- [12] E. Schaller et al., Review Article 'Electronic Noses' and Their Application to Food. 1997.
- [13] A. D. Wilson, «Applications and Advances in Electronic-Nose Technologies», p. 50, 2009.
- [14] J. C. R. Gamboa, «MODELO DE INFERENCIA DE LA RESPUESTA DE UN SENSOR DE GAS DE ESTADO SÓLIDO PARA SISTEMAS DE OLFATO ELECTRÓNICO», p. 109, 2013.
- [15] J.-E. Haugen, «ELECTRONIC NOSES IN FOOD ANALYSIS», p. 2.
- [16] M. Millán, «Diseño e implementación de un sistema de medida de gases con Arduino.», p. 126, 2016.
- [17] R. C. Pandey, M. Verma, y L. K. Sahu, «Internet of Things (IOT) Based Gas Leakage Monitoring and Alerting System with MQ-2 Sensor», vol. 5, n.o 2, p. 4, 2017.
- [18] M. S. Kasbe, S. L. Deshmukh, T. H. Mujawar, V. D. Bachuwar, L. P. Deshmukh, y A. D. Shaligram, «An Electronic nose with LabVIEW using SnO₂ Based Gas Sensors: Application to test freshness of the fruits», vol. 6, n.o 4, p. 7, 2015.
- [19] K. S. Babu, «Calibration of MQ-7 and Detection of Hazardous Carbon Mono-oxide Concentration in Test Canister», p. 7, 2018.
- [20] N. Telagam, N. Kandasamy, N. Prasad G, y M. Nanjundan, «Smart Sensor Network Based High Quality Air Pollution Monitoring System Using Labview», *Int. J. Online Eng. IJOE*, vol. 13, n.o 08, p. 79, ago. 2017, doi: 10.3991/ijoe.v13i08.7161.
- [21] «MQ135 Air Quality Sensor Module», B.M. Embedded Solutions. <https://bm-es.com/product/mq135-sensor-module/> (accedido oct. 30, 2019).
- [22] C. E. C. Peñaloza y A. J. J. Alfaro, «Nariz electrónica: Herramienta para detección de gases empleando redes neuronales artificiales.», p. 10, 2018.
- [23] G. B. Nagua, «Implementación de interfaz gráfica para la clasificación y reconocimiento de rostros mediante el procesamiento digital de señales», p. 121, 2019.
- [24] F. Pedregosa et al., «Scikit-learn: Machine Learning in Python», *Mach. Learn. PYTHON*, p. 6.
- [25] R. P. Vaz, «DESARROLLO DE UN SISTEMA DE ANÁLISIS DE DATOS EN LA INDUSTRIA DE LA AUTOMOCIÓN», p. 75.
- [26] H. Goonewardana, «PCA: Application in Machine Learning», *Medium*, feb. 28, 2019. <https://medium.com/apprentice-journal/pca-application-in-machine-learning-4827c07a61db> (accedido may 05, 2020).
- [27] Rambal, «Sensores de gas serie MQ», 2020. <https://rambal.com/gases-polucion/1254-kit-sensores-de-gas-serie-mq.html> (accedido abr. 12, 2020).
- [28] A. F. Flórez, A. C. Durán, y . Jennifer Carrillo, «Equipo de medición para la detección y clasificación de volátiles en el proceso de fermentación de la almendra de cacao en Cúcuta – norte de Santander», Tesis, Pamplona, Norte de Santander, 2019.
- [29] I. C. Pérez, «El lenguaje de programación Python/The programming language Python», *Cienc. Holguín*, p. 14, 2014.
- [30] administrar, «Cacao colombiano, nuevamente entre los mejores del mundo en París», SAC - Sociedad de Agricultores de Colombia, ago. 05, 2019. <https://sac.org.co/cacao-colombiano-nuevamente-entre-los-mejores-del-mundo-en-paris/> (accedido may 04, 2020).