HERRAMIENTAS DE CIENCIA DE DATOS APLICADAS EN LA MATERIA DE SISTEMAS PROGRAMABLES

DATA SCIENCE TOOLS APPLIED IN THE COURSE OF PROGRAMMABLE SYSTEMS

Carlos Federico Hernández Farfán
Tecnológico Nacional de México / ITS de Irapuato, México
carlos.hf@irapuato.tecnm.mx

Javier Silvestre Zavala
Tecnológico Nacional de México / ITS de Irapuato, México
javier.sz1@irapuato.tecnm.mx

Fabiola Margarita Campos Guevara
Tecnológico Nacional de México / ITS de Irapuato, México
LIS21110197@irapuato.tecnm.mx

María de Jesús Santoyo Sánchez
Tecnológico Nacional de México / ITS de Irapuato, México
LIS21111027@irapuato.tecnm.mx

José Emilio Gutiérrez Sánchez
Tecnológico Nacional de México / ITS de Irapuato, México
LIS21110797@irapuato.tecnm.mx

Recepción: 8/diciembre/2024 **Aceptación:** 7/marzo/2025

Resumen

En el curso de la materia de Sistemas Programables, se incluyen temas tales como sensores, actuadores, microcontroladores e interfaces hombre-máquina, máquina-hombre, para el monitoreo y control de procesos automatizados. En dichos procesos se pueden generar una gran cantidad de datos que pueden ser analizados para obtener información útil del proceso. Sin embargo, la obtención de datos para su almacenamiento, procesamiento y modelado para obtener información, no están incluidos en el temario actual de la materia; por lo que en este trabajo se plantea la forma en que se pueden aplicar herramientas de ciencia de datos en la materia. Se presenta la forma de obtener de datos de los sensores y actuadores para su manejo informático, su almacenamiento, procesamiento estadístico, visualización y modelado.

Palabras Clave: Ciencia de datos, Ingeniería en sistemas computacionales, Sistemas programables.

Abstract

In the study of Programmable Systems course, topics such as sensors, actuators, microcontrollers and human-machine, machine-human interfaces are included, for the monitoring and control of automated processes. In these processes, a large amount of data can be generated that can be analyzed to obtain useful information about the process. However, obtaining data for storage, processing and modeling, to obtain information, is not included in the current syllabus of the subject; therefore, this work proposes how data science techniques can be applied to the subject. The way to obtain data from sensors and actuators for computer management, storage, statistical processing, visualization and modeling is presented.

Keywords**:** *Computer sytems engineering, Data science, Programmable systems.*

1. Introducción

La ciencia de datos, la industria 4.0 y el internet de las cosas, son tecnologías emergentes que se están consolidando rápidamente por su impacto en los sectores productivos y de servicios. Dichas tecnologías son afines a las carreras de ingeniería, sin embargo, su asimilación o inclusión en los programas de estudio, no ha sido tan rápido como su desarrollo. Por ejemplo, en el artículo de [Hicks, 2018], se menciona que la demanda de capacitación en ciencia de datos está aumentando y se aboga por una actualización en los planes de estudio.

La ciencia de datos, que incluye áreas de estudio tales como la estadística, las bases de datos, la programación, el aprendizaje automático y alguna disciplina específica para su aplicación; la industria 4.0, que incluye el manejo de grandes cantidades de datos para el monitoreo y control de procesos industriales; y el internet de las cosas, que incluye la aplicación de las tecnologías de las redes de computadoras e informática en dispositivos de la vida cotidiana; son todas áreas y conceptos afines a la carrera de Ing. en Sistemas computacionales. Sin embargo, dichos temas y conceptos no se encuentran integrados de manera explícita en los

temarios de las materias como parte del estudio de ciencia de datos, industria 4.0 o internet de las cosas.

Dentro de la carrera de Ingeniería en Sistemas Computacionales del Tecnológico Nacional de México campus Instituto Tecnológico Superior de Irapuato, se imparte la materia de Sistemas Programables. La principal competencia a desarrollar en el curso consiste en realizar interfaces hombre – máquina para el monitoreo y control de procesos automatizados. Los temas que se incluyen en el curso son: sensores, actuadores, microcontrolador, programación del microcontrolador, comunicación del microcontrolador, y diseño de interfaces hombre - máquina, tal como se muestra en [Tecnológico Nacional de México campus Instituto Tecnológico Superior de Irapuato, 2024]. En los procesos de automatización industrial, se genera una gran cantidad de datos del proceso, que se pueden analizar para obtener información útil, por lo que se puede decir que la materia de Sistemas Programables es afín a los conceptos de ciencia de datos.

Cuando se imparte la materia es común que se presenten algunas situaciones problemáticas adicionales tales como recursos limitados de infraestructura de laboratorio, la búsqueda de un balance de contenido teórico y práctico del curso; y en este caso, la búsqueda de un enfoque introductorio del curso hacia la ciencia de datos.

En este trabajo se propone mostrar la viabilidad de cómo pueden ser incluidas algunas herramientas de ciencia de datos en las prácticas de la materia de Sistemas Programables, tratando de no sobrecargar de actividades adicionales a los estudiantes, sino buscar dar un enfoque al curso, más hacia las ciencias computacionales y hacia un enfoque introductorio a la ciencia de datos, su relación y aplicación en el curso; y de acuerdo a los recursos de infraestructura de laboratorio disponible.

En atención a la situación problemática mencionada y al objetivo de este trabajo, en lo que se refiere a la revisión bibliográfica, tanto en el libro de Sistemas Programables avanzados de [Espinosa, 2017], como en el libro de Sistemas Programables Avanzados de [Mercado, 2019], se incluye un amplio estudio de programación de un controlador lógico programable (PLC), empleado en la

automatización industrial; sin embargo no incluyen aspectos relacionados con el tratamiento informático de los datos generados en el proceso de automatización, además de que se enfocan en el uso de un PLC de fabricante en particular.

Continuando con la revisión bibliográfica, para identificar algunas de las herramientas usadas comúnmente en ciencia de datos, se revisó el artículo Ciencia de Datos – una revisión del estado del arte de [Méndez, 2018], el libro de texto Ciencia de Datos - Teoría y Aplicaciones de [Ferreira, 2022], el artículo Teaching Data Science de [Brunner, 2016] de un curso introductorio de ciencia de datos, el artículo Data Science with Excel de [Setiyanto, 2022]; así como el contenido temático del Diplomado en Ciencia de Datos de la [Universidad Nacional Autónoma de México, 2024], como del Diplomado en Ciencia de Datos del [Tecnológico Nacional de México, 2024].

En el artículo de [Méndez, 2018], se menciona que las técnicas más utilizadas en ciencia de datos son: regresión lineal, k-nearest neighbors, k-means, regresión logística, redes bayesianas, máquina de soporte vectorial y redes neuronales. En el artículo de [Brunner, 2016], se presenta la estructura de un curso introductorio a la ciencia de datos, el cual incluye temas como: sistema operativo Unix y manejo de interfaz de línea de comandos; programación en lenguaje Python y uso de librerías como Numpy, Matplotlib y Seaborn para la visualización de datos, la librería Pandas para el manejo estadístico de los datos; también se incluye el tema de bases de datos y uso del lenguaje SQL, y finalmente una introducción a machine learning, en particular el uso de los métodos k-nearest neighbors, support vector machine, decision trees, principal components analysis y k-means.

En el artículo de [Setiyanto, 2022], se muestra la forma en que se puede emplear Excel en ciencia de datos. El libro de Ciencia de Datos de [Ferreira, 2022], incluye temas como manejo de archivos CSV, almacenamiento en bases de datos, estadística, aprendizaje automático con algoritmos de regresión lineal y agrupamiento k-means, deep learning y procesamiento en lenguaje natural; empleando el lenguaje de programación Python y librerías como Pandas, Matplotlib, etc. También se revisó el contenido del Diplomado en Ciencia de Datos de la Universidad Nacional Autónoma de México, el cual incluye temas como estadística,

manipulación y visualización de datos con Python, machine learning, deep learning, procesamiento de lenguaje natural, entre otros como se muestra en [Universidad Nacional Autónoma de México, 2024]. En el diplomado en Ciencia de Datos del Tecnológico Nacional de México, se incluyen los temas: estadística, bases de datos y lenguaje SQL, programación en Python y uso de las librerías Pandas, Numpy, Matplotlib; también se incluye el tema de machine learning con el uso de la librería Scikit-Learn para realizar el algoritmo de regresión lineal y el algoritmo k-means para realizar agrupamiento o clustering, tal como se muestra en [Tecnológico Nacional de México, 2024]. De esta manera, se identificaron y seleccionaron las siguientes herramientas básicas y que son comunes en ciencia de datos, para aplicarse en el curso de Sistemas Programables: el uso de archivos en formato CSV compatibles con hojas de cálculo y bases de datos para el manejo de los datos provenientes de los sensores y actuadores; el uso de estadística básica descriptiva para obtener información de los datos de los sensores y actuadores; el uso de diferentes gráficos de visualización de datos; el uso del software Excel. También se identificó el uso y almacenamiento de información en bases de datos mediante el lenguaje SQL (Structured Query Language). Se identificó el uso del lenguaje de programación Python y el uso de diferentes librerías para el manejo de datos. Se identificó el uso de algoritmos aprendizaje automático o machinne learning, para realizar predicciones del comportamiento de una variable mediante regresión lineal, y el uso del algoritmo k-means de agrupamiento de datos o clustering.

Continuando con la revisión bibliográfica, en lo que se refiere a herramientas didácticas para la materia, en un trabajo previo de [Hernández, 2018], se realizó una propuesta didáctica al mostrar cómo se podría incorporar al curso de Sistemas Programables, el uso de la computadora de bajo costo Raspberry Pi. La computadora Raspberry Pi, cuenta con un sistema operativo de una distribución de Linux, con entornos de programación en Python, que permite el manejo de sensores y actuadores, así como la generación de archivos CSV; permite el uso de librerías para la elaboración de interfaces hombre-máquina y la ejecución de librerías empleadas en ciencia de datos; también permite la instalación de bases de datos, comunicaciones en red y servicio web.

2. Métodos

Se emplearon algunas herramientas de ciencia de datos en las prácticas del curso de la materia de Sistemas Programables de la carrera de ingeniería en Sistemas Computacionales del Tecnológico Nacional de México campus Instituto Tecnológico Superior de Irapuato, con un grupo de 30 estudiantes. El curso se impartió en el periodo de verano 2024 con una duración de $75\,horas$, con una distribución de tiempo de 50% teórico y 50% práctico.

En el curso se realizaron 20 prácticas para cubrir el contenido temático de la materia. Se emplearon los sensores y actuadores indicados en el temario de la materia, entre los cuales se incluye el sensor óptico CNY70, el sensor de temperatura LM35 y el sensor de proximidad HC-SR04; y como dispositivos actuadores, un motor corriente directa de $5\,V$, un servomotor SG-90 y un motor a pasos 28BYJ-48 de $5\,V$. Se empleó el simulador de circuitos electrónicos en línea Tinkercad de la compañía [Autodesk, 2024], para la simulación de sensores y actuadores mediante la tarjeta Arduino. Para la adquisición de datos de los sensores y manejo de los actuadores, se empleó la tarjeta electrónica basada en microcontrolador Arduino Uno en las prácticas iniciales y la computadora de bajo costo Raspberry Pi 3 Modelo B en el resto de las prácticas.

Para identificar y seleccionar las herramientas comúnmente utilizadas en ciencia de datos y aplicarlas en el curso, se realizó una revisión bibliográfica como se describe en la sección anterior.

Los datos de los sensores y actuadores se registraron en un archivo tipo CSV, que permite obtener los datos en forma de tabla compatible con hojas de cálculo y bases de datos. Los datos obtenidos de los sensores y actuadores, se registraron en una hoja de cálculo Excel 2016, para el manejo estadístico básico, visualización gráfica y obtención de un modelo estadístico de los datos. Mediante un archivo CSV, los datos se importaron y se almacenaron en una base de datos para realizar consultas de información de dichos datos mediante el lenguaje SQL. Se empleó el gestor de bases de datos MariaDB 10.10.2 y la interfaz gráfica para el manejo de la base de datos HeidiSQL 10.3.0. Se empleó el lenguaje de programación Python en el entorno de programación IDLE 3.11.1 para el manejo de sensores y actuadores, así

como para el manejo estadístico básico, visualización gráfica y obtención de un modelo estadístico de los datos.

Se empleó el software Excel y/o el lenguaje de programación Python, para obtener la estadística descriptiva de los datos, tales como valores máximos y mínimos, medidas de tendencia central (media, mediana, moda), medidas de dispersión (rango, varianza, desviación estándar), medidas de posición (deciles, cuartiles, percentiles), y distribución de probabilidad normal. Se realizaron diferentes gráficas de visualización de datos, como histogramas, columnas agrupadas, gráficas de dispersión, distribución de Gauss y diagramas de caja. Finalmente se obtuvo el modelo o patrón característico del comportamiento de los datos para realizar predicciones y/o agrupamiento de datos.

Se aplicaron algoritmos de aprendizaje automático para obtener un modelo de comportamiento de los datos de los sensores, mediante lenguaje de Programación Python y empleando la librería Scikit-Learn; disponible a través del sitio web de [Scikit-Learn, 2024] y descrita en el artículo de [Pedregosa, 2011]. Se ejecutó un algoritmo de aprendizaje automático supervisado, que permite obtener un modelo estadístico de los datos conocidos para realizar alguna predicción sobre el comportamiento de un conjunto de datos nuevos y se ejecutó un algoritmo aprendizaje automático no supervisado de agrupamiento de datos de un sensor.

3. Resultados

En esta sección se presentan algunos ejemplos del uso de herramientas de Ciencia de Datos en prácticas de laboratorio del curso de Sistemas Programables.

Mediante el sensor de temperatura LM35 y la tarjeta Ardunio Uno, se realizó la medición de temperatura ambiente, los datos se obtuvieron mediante muestras cada $0.1\ s$ durante $5\ min$, obteniéndose 3000 datos aproximadamente y registrándose en un archivo CSV. La Tabla 1muestra una parte de los datos registrados en el archivo CSV y la Tabla 2 los datos estadísticos como cantidad total de datos, promedio, mediana, moda, valor máximo, mínimo, varianza y desviación estándar, obtenidos mediante una hoja de cálculo en Excel, Figura 1, el histograma de frecuencia de los datos y la gráfica de distribución de Gauss.

Tabla 1 Temperatura del sensor LM35.

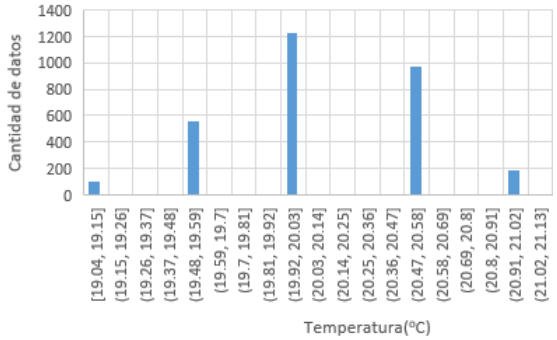
Número	Temperatura ($°C$)
1	20.51
2	20.02
3	20.51
4	19.04
5	20.51
6	20.51
7	19.53

Fuente: elaboración propia

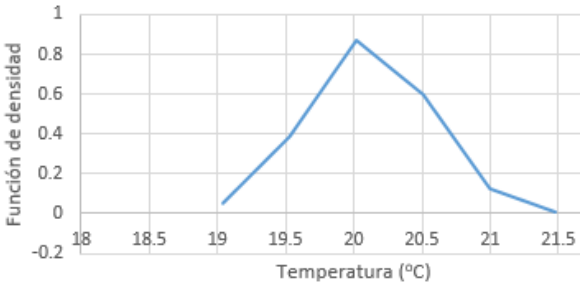
Tabla 2 Estadísticas del sensor LM35.

Datos	3057
Promedio	20.10 $°C$
Mediana	20.02 $°C$
Moda	20.02 $°C$
Máximo	24.48 $°C$
Mínimo	19.04 $°C$
Varianza	0.20
Desviación estándar	0.58

Fuente: elaboración propia



a) Histograma de las mediciones de temperatura del sensor LM35.



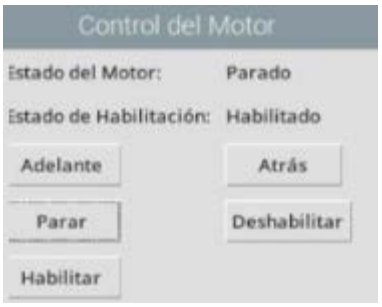
b) Distribución normal de Gauss de las mediciones de temperatura del sensor LM35.

Fuente: elaboración propia

Figura 1 Uso del sensor de temperatura LM35.

Mediante la computadora de bajo costo Raspberry Pi, se realizó el control de la dirección de giro de un motor de corriente directa de $5\ V$, el programa para el control

del motor y la interfaz de usuario, se realizó en lenguaje de programación Python, y se realizó el registro de las acciones sobre el motor en un archivo CSV como su estado de parado o detenido, la dirección hacia adelante o hacia atrás, y fecha y hora del registro en el archivo CSV. En la Figura 2, se muestra la interfaz para el manejo del motor.



Fuente: elaboración propia

Figura 2 Control de un motor de corriente directa.

Mediante el uso del sensor ultrasónico de proximidad HC-SR04 y la tarjeta Arduino, se obtuvieron datos correspondientes a mediciones de distancia. Se realizaron mediciones a las distancias de $90, 75, 60, 45, 30$ y 15 cm, para cada distancia se obtuvieron muestras cada 0.1 s con 1 min de duración cada una obteniéndose en total 3600 datos aproximadamente. Los datos se almacenaron en un archivo tipo CSV y posteriormente fue importado en la base de datos MariaDB. En la Figura 3, se muestran las sentencias en lenguaje SQL para realizar diferentes consultas para obtener información estadística sobre los datos.

```
SELECT Distancia AS 'Distancia (cm)' FROM sensor_hcsr04;
SELECT COUNT(*) AS 'Total datos'FROM sensor_hcsr04;

SELECT MAX(Distancia) FROM sensor_hcsr04;
SELECT MIN(Distancia) FROM sensor_hcsr04;
SELECT ROUND(AVG(Distancia)) FROM sensor_hcsr04;

SELECT Distancia AS 'Datos a 90 cm' FROM sensor_hcsr04
WHERE Distancia < 95 AND Distancia > 85;
SELECT COUNT(*) AS 'Total datos a 90 cm' FROM sensor_hcsr04
WHERE Distancia < 95 AND Distancia > 85;

SELECT Distancia AS 'Datos a 75 cm' FROM sensor_hcsr04
WHERE Distancia < 80 AND Distancia > 70;
SELECT COUNT(*) AS 'Total datos a 75 cm'FROM sensor_hcsr04
WHERE Distancia < 80 AND Distancia > 70;
```

Fuente: elaboración propia

Figura 3 Consultas SQL en base de datos del sensor HC-SR04.

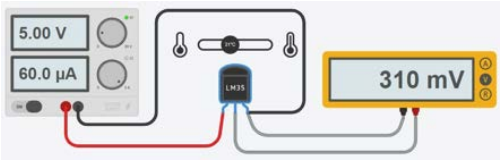
En la Tabla 3, se muestra el resultado de las consultas SQL como el listado de los datos almacenados en una tabla de la base de datos, la cantidad total de datos, el valor máximo, mínimo y promedio, el listado y cantidad de datos correspondientes a una distancia en particular.

Tabla 3 Resultados de las consultas SQL.

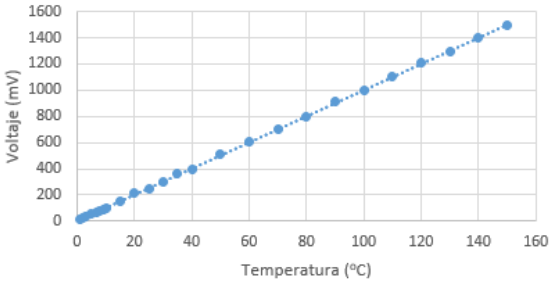
Datos	3471
Distancia máxima	106 cm
Distancia mínima	15 cm
Distancia promedio	53 cm
Datos a 90 cm	580
Datos a 75 cm	565

Fuente: elaboración propia

En la Figura 4a, se muestra la simulación en Tinkercad del sensor de temperatura $LM35$, se puede observar el ajuste de temperatura en grados centígrados y el voltaje correspondiente en mV generado por el sensor. Los datos de las mediciones del sensor son registrados de forma manual en un archivo CSV para su posterior procesamiento.



a) Simulación del sensor de temperatura LM35.



b) Gráfica de dispersión de datos con línea de tendencia.

Fuente: elaboración propia

Figura 4 Simulación y gráfica de datos del sensor $LM35$.

En total se registraron 25 $datos$ de temperatura y voltaje, en un rango de valores de 1 a 150 °C, correspondientes a un rango de 10 a 1500 mV. En la Figura 4b, se

muestra la gráfica de dispersión de datos realizada en la hoja de cálculo Excel, se puede observar a simple vista, la respuesta lineal del sensor de $10\ \frac{mV}{^{\circ}C}$, tal como se muestra en la hoja de datos del fabricante [Texas Instruments Incoporated, 2024].
En la Ecuación 1, se muestra la línea de tendencia obtenida mediante Excel que corresponde a la regresión lineal y que representa el modelo de comportamiento del sensor.

$$\mathrm{y} = 10.002\ \mathrm{t} + 2.68 \tag{1}$$

Dónde:

y : Voltaje de salida del sensor.

t : Temperatura ambiente.

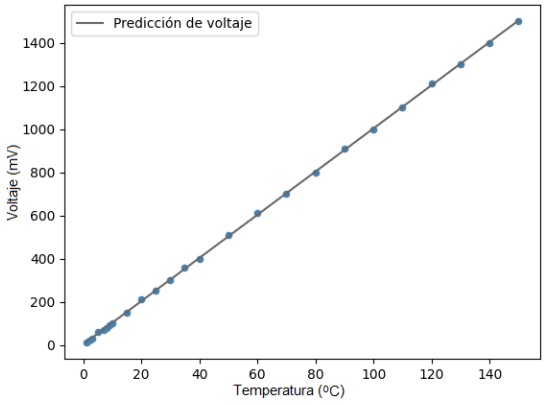
En otro ejemplo, a partir del archivo CSV de los datos del sensor de temperatura LM35, se emplearon las siguientes librerías del lenguaje de programación Python: la librería Pandas, para el procesamiento en forma de tablas de los datos; la librería Matplotlib para la visualización gráfica de los datos; y la librería Scikit-Learn de aprendizaje automático, para obtener el modelo de comportamiento de los datos mediante una regresión lineal y poder realizar predicciones acerca del comportamiento de las variables. Mediante la librería Pandas, se realiza la lectura del archivo CSV de los datos de sensor y se obtiene un marco de datos o dataframe.
En la Tabla 4, se muestran los datos correspondientes a la regresión lineal mediante la librería de aprendizaje automático Scikit-Learn de Python. A partir de la obtención del modelo de los datos, es posible realizar la predicción de los valores de voltaje correspondientes a una determinada temperatura. La Ecuación correspondiente a la regresión lineal obtenida mediante la librería Scikit-Learn es la misma que se muestra en la Ecuación 1 obtenida mediante la hoja de cálculo Excel, pero obtenida de diferente forma además de que se pueden obtener las predicciones del voltaje para una determinada temperatura. En la Figura 5, se puede observar la gráfica correspondiente al modelo obtenido mediante una regresión lineal, empleando la librería Matplotlib.
Mediante el sensor ultrasónico de distancia HC-SR04 y la tarjeta Aduino Uno, se realizaron mediciones a $90, 75, 60, 45, 30\ y\ 15\ cm$; para cada distancia se obtuvieron

datos durante $1\ min$ y cada muestra se obtuvo cada $0.1\ s$, obteniéndose un total de $3469\ datos$ y un promedio de $566\ datos$ para cada distancia. Los datos obtenidos del sensor se almacenaron en un archivo CSV para su tratamiento estadístico y visualización gráfica.

Tabla 4 Datos del sensor LM35.

Número	Temperatura ($°C$)	Voltaje (mV)	Predicción de Voltaje (mV)
0	1	10	12.68
1	2	20	22.68
2	3	30	32.68
3	5	60	52.69
4	7	70	72.69
5	8	80	82.69
6	9	90	92.70
7	10	100	102.70
8	15	150	152.71
9	20	210	202.72

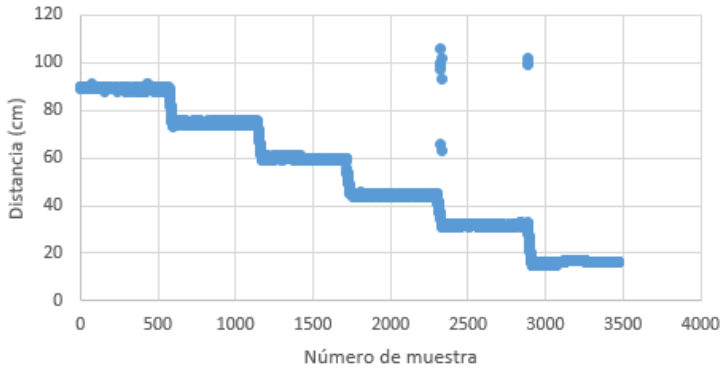
Fuente: elaboración propia



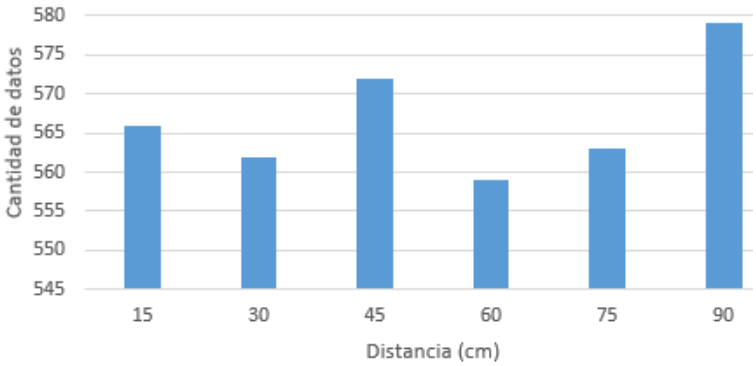
Fuente: elaboración propia

Figura 5 Regresión lineal con predicciones para el sensor LM35.

En la Figura 6a, se muestra la gráfica de dispersión de datos obtenida mediante la hoja de cálculo Excel. Se observa que los datos se pueden agrupar por distancia de medición, un grupo de aproximadamente $500\ datos$ para una distancia de $90\ cm$, otro grupo de aproximadamente $500\ datos$ para una distancia de $75\ cm$, y así sucesivamente, pudiéndose observar 6 grupos de datos. En la Figura 6b, se muestra una gráfica de columna agrupada con la cantidad de datos obtenidos para cada distancia, nuevamente se destaca que son aproximadamente $500\ datos$ por grupo o por distancia de medición.



a) Gráfica de dispersión de datos del sensor HC-SR04.



b) Cantidad de datos para cada distancia de medición.

Fuente: elaboración propia

Figura 6 Mediciones del sensor ultrasónico HC-SR04.

A partir del archivo CSV de los datos del sensor de proximidad HC-SR04, se emplearon las siguientes librerías del lenguaje de programación Python: la librería Pandas, para el procesamiento en forma de tablas de los datos; la librería Matplotlib para la visualización gráfica de los datos; y la librería Scikit-Learn de aprendizaje automático, para realizar un agrupamiento de los datos mediante el método K-means (k grupos de pertenencia al promedio cercano).

En la Tabla 5, se muestra la estadística básica de los datos del archivo del sensor HC-SR04 como cantidad de datos total, promedio, valor mínimo, valor máximo, desviación estándar, primer, segundo y tercer cuartil. En la Tabla 6, se muestra el resultado del agrupamiento de valores del sensor de acuerdo a las diferentes mediciones de distancia realizadas y obtenido por medio del algoritmo k-means. Se obtuvo un agrupamiento de 3471 datos en 6 grupos o clusters, y en la Figura 7 se muestra la gráfica del resultado del agrupamiento, se puede observar que el agrupamiento corresponde a las 6 diferentes mediciones de distancia realizadas.

Tabla 5 Datos del sensor ultrasónico.

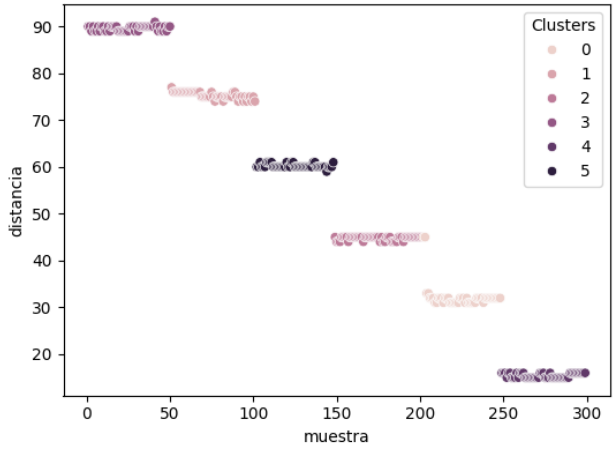
	Muestra	**Distancia**
Cantidad	3471	3471
Media	1736	52.86 cm
Std	1002.13	24.97 cm
Mínimo	1.00	15.00 cm
25%	868.50	32.00 cm
50%	1736.00	52.00 cm
75%	2603.50	75.00 cm
Máximo	3471	106.00 cm

Fuente: elaboración propia

Tabla 6 Agrupamiento de datos.

Número	**Muestra**	**Distancia (cm)**	**Clúster**
0	1	90	3
1	2	89	3
2	3	90	3
3	4	89	3
…	…	…	…
3466	3467	16	5
3467	3468	16	5
3468	3469	16	5

Fuente: elaboración propia



Fuente: elaboración propia

Figura 7 Agrupamiento de datos del sensor ultrasónico.

A partir del análisis de los datos de los sensores, es posible identificar valores que puedan indicar de forma temprana el posible fallo en algún proceso, tal como se muestra en el artículo IoT System for Mechanical Faults detection in automotive industry de [Ramírez, 2023].

4. Discusión

Las actividades de aprendizaje del curso pueden haberse enfocado hacia diferentes áreas de especialización tales como a los sistemas de adquisición de datos, supervisión y control, como los sistemas SCADA (Supervisory Control And Data Acquisition); o bien hacia la programación de controladores lógicos programables (PLC), o por ejemplo hacia los sistemas de bus de comunicación entre sensores y el sistema de control, como Profibus, etc. Sin embargo, el enfoque en este trabajo, es hacia un área más afín a la carrera de ingeniería en sistemas computacionales, como lo es campo de la ciencia de datos, la cual incluye el uso del lenguaje de programación Python, el uso de bases de datos y el uso de la inteligencia artificial; y de acuerdo a la infraestructura de laboratorio disponible.

El trabajo realizado tiene un enfoque didáctico y no se trata de un curso de ciencia de ciencia de datos, sino de dar un enfoque introductorio que motive su estudio posterior al aplicar algunas de sus herramientas en el curso de Sistemas Programables.

El desarrollo de este trabajo tiene un enfoque que permita mostrar un panorama introductorio de las herramientas empleadas comúnmente en ciencia de datos. Existe una gran variedad de herramientas estadísticas, como se muestra por ejemplo en [Navidi, 2022], así como una gran variedad de herramientas, técnicas y aplicaciones de ciencia de datos, sin embargo, sobrepasan el objetivo y el alcance de este trabajo.

5. Conclusiones

Se presentó una forma en la que, mediante el uso de la estadística, la programación, el manejo de bases de datos y el aprendizaje automático, en un curso de Sistemas Programables, es viable dar un enfoque introductorio del curso hacia la ciencia de datos. Con este trabajo se contribuye a mejorar el balance teórico y práctico de los conceptos del curso y su implementación física. Se implementaron, además de las prácticas normales del curso, algunas herramientas comúnmente utilizadas en ciencia de datos, tales como medidas de tendencia central, medidas de dispersión, gráficos de visualización de datos, programación en lenguaje Python,

y el uso de las librerías Matplotlib, Pandas y Numpy; así como el uso de las librerías Scikit-Learn de aprendizaje automático. Se mostró de forma didáctica como se pueden incluir dichas herramientas en el curso, buscando apegarse completamente al contenido temático del curso sin sobrecargar de actividades a los estudiantes, dando el enfoque hacia un área afín a su especialidad, utilizando la menor cantidad de infraestructura de laboratorio y costos de material para el desarrollo de las prácticas.

6. Bibliografía y Referencias

[1] Autodesk. (24 de febrero de 2024). Tinkercad. https://www.tinkercad.com/.

[2] Brunner, R., Kim, E. (2016). Teaching Data Science. Procedia Computer Science, Vol.80, 1947-1956, DOI: 10.1016/j.procs,2016.05.513, ISSN 1877-0509.

[3] Espinosa, J. (2017). Sistemas Programables Avanzados. Alfaomega Marcombo, ISBN: 978-607-622-732-9.

[4] Ferreira, R. (2022). Ciencia de datos – Teoría y Aplicaciones. Tecnológico Nacional de México.

[5] Hernández, C., Márquez, A., Ávila, D. (2018). Apoyo didáctico de la tarjeta Raspberry Pi en la material de Sistemas Programables. Pistas Educativas, Vol. 40 (No. 130), 2185-2196, ISSN:2448-847X.

[6] Hicks, S., Irizarry, R. (2018). A Guide to Teaching Data Science. Am Stat, 72(4):382-39, doi: 10.1080/00031305.2017.1356747.

[7] Méndez, N., Rubier, J. (2018). Ciencia de datos: una revisión del estado del arte. UCE Ciencia Revista de Posgrado, Vol. 6 (No.3).

[8] Mercado, J. Sistemas Programables Avanzados. Paraninfo, ISBN: 978 842 834 2292.

[9] Navidi, W. (2022). Estadística para ingenieros y científicos. Mc Graw Hill, 5a Edición, ISBN: 978607-15-1757-9.

[10] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Weiss, R. Dubourg, V. Vanderplas, J., Passos, A., Cournapeau,

D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-Learn: Machine Learning in Python. Journal of Machine Learning Research, Vol.12.

[11] Ramírez, C., García, J., Serrano, J. (2023). IoT System for mechanical faults detection in automotive industry. 2023 IEEE International Autumn Meeting on power, electronics and computing, 1-6, DOI: 10.1109/ROPEC58757.2023.10409331.

[12] Scikit-Learn. (2024). (2024). Machine Learning in Python. Obtenido de https://scikit-learn.org/stable/.

[13] Setiyanto, S., Setiawan, I. (2022). Data Science with Excel. International Journal of Computer and Information System, Vol 3. (No.3), e-ISSN: 2745-9659.

[14] Tecnológico Nacional de México. (2024). Diplomado en Ciencia de Datos. https://cienciadedatos.foprode. aguascalientes.tecnm.mx/ index.html.

[15] Tecnológico Nacional de México campus Instituto Tecnológico Superior de Irapuato. (2023). Ingeniería en Sistemas Computacionales. Oferta Educativa, https://irapuato.tecnm.mx/moferta/sistcomputacionales/pdf /plan_estudios/7%20Sistemas%20Programables.pdf.

[16] Texas Instruments Incorporated. (2024). LM35 Precision Centigrade Temperature Sensors. https://www.ti.com/lit/ds/symlink/lm35.pdf?ts=1719308250639&ref_url=https%253A%252F%252Fwww.google.com%252F.

[17] Universidad Nacional Autónoma de México. (20 de Junio de 2024). Diplomado en Ciencia de Datos. https://docencia.tic.unam.mx/ diplomado-en-ciencia-de-datos.