

Universidad de Ingeniería y Tecnología

TESIS DE PREGRADO

Propuesta de gestión para la detección de fallas en colectores de aguas residuales del distrito de San Juan De Lurigancho

Reyes Diaz, Alberto Franco

Award date:
2021

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA
CARRERA DE INGENIERÍA MECÁNICA



**PROPUESTA DE GESTIÓN PARA LA DETECCIÓN
DE FALLAS EN COLECTORES DE AGUAS
RESIDUALES DEL DISTRITO DE SAN JUAN DE
LURIGANCHO**

TESIS

Para optar el título profesional de Ingeniero Mecánico

AUTOR

Alberto Franco Reyes Díaz (ORCID: 0000-0001-6223-2382)

ASESOR

Elmer Ramirez (ORCID: 0000-0003-4089-7871)

Lima – Perú

2021

La tesis

Propuesta de gestión para la detección de fallas en colectores de aguas residuales del distrito de San Juan de Lurigancho

Ha sido aprobada

[Nombres y apellidos del Presidente de Jurado]

[Elmer Ramirez]

[Nombres y apellidos del Tercer jurado]

Dedicatoria:

A Dios, por brindarle los recursos a mis padres y contar con su favor para que mis estudios universitarios fueran posibles y llegar a concluir esta etapa de mi vida de profesional con los conocimientos necesarios para mi desarrollo, aporte a la sociedad y a mi familia. A mis padres por su soporte, motivación, disciplina y recomendaciones para que yo sea un buen profesional.

A mi madre, una mujer devota de Dios, que, con todo su amor y cariño, me apoyó en los cursos que trataban sobre economía y me exigió que diera lo mejor de mí en todos los cursos siendo a la vez paciente conmigo.

A mi padre, otro devoto de Dios, que, con mucho amor y cariño también me apoyó en los cursos que trataban de economía y contabilidad, quien también fue exigente conmigo motivándome a esforzarme en todas las materias de mi carrera.

Agradecimientos:

A Dios, quien me permitió seguir y concluir la carrera de ingeniería mecánica, ya que sin su ayuda no hubiera podido llegar a la meta y concluir de manera satisfactoria.

A mis padres que me apoyaron en todo en mis estudios a pesar de cualquier dificultad que se haya presentado a lo largo de este tiempo.

A mi asesor, el profesor Elmer Ramirez, quien me apoyó en la investigación y desarrollo de esta tesis siendo fundamental el soporte recibido.

A la Universidad de Ciencias y Tecnología UTEC, institución prestigiosa que, con sus altos estándares de enseñanza, me brindó los conocimientos necesarios para la conclusión de la carrera de ingeniería mecánica con la confianza de haber recibido una excelente formación profesional.

TABLA DE CONTENIDO

	Pág.
RESUMEN	12
ABSTRACT	13
INTRODUCCIÓN	14
CAPÍTULO I MARCO TEÓRICO	19
1.1 Sistema de distribución de Aguas servidas	19
1.1.1 Componentes	20
1.1.2 Topología	21
1.1.3 Protocolos de instalación y pruebas	23
1.1.3.1 Protocolos de instalación	24
1.1.3.2 Protocolos de pruebas	27
1.1.4 Materiales en tuberías	28
1.2 Servicio de alcantarillado en San Juan de Lurigancho	29
1.3 Fallas en sistemas de distribución	31
1.3.1 Causas de daños en el sistema de alcantarillado	31
1.4 Herramientas de diagnóstico para la detección de fallas	33
1.4.1 Inspección por medio de circuito cerrado de televisión	33
1.4.1.1 Equipos necesarios para la inspección	34
1.4.2 Otros métodos para la inspección de la condición interna de una tubería	35
1.4.2.1 Sistema de escaneo por láser	35
1.4.2.2 Sistema de escaneo por Sonar	35
1.5 Herramientas de diagnóstico por base de datos	36
1.5.1 Ventajas y Desventajas	37

1.6	Método de Análisis de falla	38
1.6.1	Machine Learning.....	38
1.6.2	Clusterización, K-mean y las 4 regresiones	39
1.6.2.1	Clusterización	39
1.6.2.2	Método de Elbow.....	44
1.6.2.3	Las 4 regresiones	45
1.6.2.4	Matriz de confusión	47
1.7	Teorema de Manning.....	49
CAPÍTULO II METODOLOGÍA		51
2.1	Establecimiento de variables	51
2.2	Selección de tramos.....	54
2.3	Obtención de datos	54
2.4	Análisis por medio de Machine Learning	55
2.5	Resultados	63
CAPÍTULO III RESULTADOS		64
3.1	Variables establecidas	64
3.1.1	Selección de tramo.....	65
3.2	Machine Learning, Clusterización y Los 4 Algoritmos para el Análisis	65
3.2.1	Condición del colector según el diámetro, edad, profundidad y pendiente	65
3.2.2	Condición según el diámetro, tirante, ángulo, área mojada, perímetro mojado, radio hidráulico, velocidad, velocidad crítica, caudal, pendiente, número de actividades y material.	70
CONCLUSIONES.....		76
REFERENCIAS BIBLIOGRÁFICAS		76
ANEXOS		81

ÍNDICE DE FIGURAS

	Pág.
Figura 1.1 Topología perpendicular	21
Figura 1.2 Topología radial	21
Figura 1.3 Topología de Interceptores	22
Figura 1.4 Topología de Abanico	22
Figura 1.5 Plano del sistema de alcantarillado de SJL	23
Figura 1.6 Población de San Juan de Lurigancho	30
Figura 1.7 Cámaras montadas sobre transportadores	34
Figura 1. 8 Comparación de diámetro teórico con diámetro original	36
Figura 1.9 Ejemplo de gráfica obtenido por el método de Elbow por medio del programa R	44
Figura 1. 10 Matriz de confusión	47
Figura 1. 11 Sección de tubo por el que pasa el flujo del agua	49
Figura 2. 1 Diagrama de metodología	51
Figura 3. 1 Gráfica de método de Elbow del Escenario 1.....	66
Figura 3. 2 Gráfica de Clusterización del Escenario 1.....	67
Figura 3. 3 Resultados de la clusterización de k-mean del Escenario 1.....	67
Figura 3. 4 Gráfica de la correlación de datos del Escenario 1.....	68
Figura 3. 5 Gráfica de método de Elbow del Escenario 2.....	71
Figura 3. 6 Gráfica de Clusterización del Escenario 2.....	72
Figura 3. 7 Resultados de la clusterización de k-mean del escenario 2.....	73
Figura 3. 8 Gráfica de Correlación del Escenario 2.....	73

ÍNDICE DE TABLAS

	Pág.
Tabla 2. 1 Tabla de definición de las variables junto con sus unidades y fuentes.....	54
Tabla 3. 1 Tabla de precisión del Escenario 1 de Regresión Logística.	68
Tabla 3. 2 Tabla de precisión de la Regresión K-NN del Escenario 1.	69
Tabla 3. 3 Tabla de precisión de la Regresión SVM del Escenario 1.	69
Tabla 3. 4 Tabla de precisión de la Regresión Decision Tree del Escenario 1.....	69
Tabla 3. 5 Ratios de precisión de cada uno de los análisis del Escenario 1.	70
Tabla 3. 6 Matriz de Confusión del análisis de Regresión Logística del Escenario 2.	74
Tabla 3. 7 Matriz de Confusión del análisis de Regresión K-NN del Escenario 2.....	74
Tabla 3. 8 Matriz de Confusión del análisis de Regresión SVM del Escenario 2.	74
Tabla 3. 9 Matriz de Confusión del análisis de Regresión Decision Tree del Escenario 2.	75
Tabla 3. 10 Ratios de cada uno de los análisis del Escenario 2.	75

ÍNDICE DE ANEXOS

	Pág.
ANEXO 1: PRODUCCIÓN DE AGUA POTABLE Y TRATAMIENTO DE AGUAS SERVIDAS DEL 2018.....	82
ANEXO 2: TABLA DE ÍNDICE DE INCIDENCIAS EN SJL SEGÚN SEDAPAL	84
ANEXO 3: Cuadros de los escenarios 1 y 2.....	85
ANEXO 4: Gráficas de cada una de los regresiones del escenario 1	92
ANEXO 5: Gráficas de cada una de los regresiones del Escenario 2	96

ÍNDICE DE ECUACIONES

	Pág.
Ecuación 1.1 Ecuación para la exactitud	48
Ecuación 1.2 Ecuación para calcular la exactitud promedio	48
Ecuación 1.3 Ecuación para la sensibilidad	48
Ecuación 1.4 Ecuación para la sensibilidad promedio	48
Ecuación 1.5 Ecuación para la precisión	49
Ecuación 1.6 Ecuación para la precisión promedio	49
Ecuación 1.7 Ecuación para el tirante	50
Ecuación 1.8 Área mojada del ducto	50
Ecuación 1.9 Radio hidráulico del área mojada	50
Ecuación 1.10 Perímetro que ocupa el área mojada	50
Ecuación 1.11 Caudal del área mojada	50
Ecuación 1.12 Velocidad crítica	50

RESUMEN

En la presente tesis se evaluaron tres escenarios de fallas en colectores de aguas residuales del distrito de San Juan de Lurigancho. El primer análisis consistió en evaluar la condición de los alcantarillados de acuerdo al diámetro, la pendiente, el tiempo (en años), el número de fallas acumulado y la profundidad, y el segundo análisis consistió en evaluar si un colector sufre un atoro de acuerdo al número de actividades usando como variables el diámetro, la pendiente, la velocidad, el tirante, el caudal, el ángulo, el perímetro mojado, el área, el tiempo (en meses) y el radio hidráulico. El análisis se realizó por medio del método conocido como Machine Learning, que sirvió para realizar predicciones en los dos escenarios por medio del método de clusterización de K-means para saber qué dice la gráfica de clusterización para cada escenario y después aplicar las 4 regresiones las cuales son la logística, la de K-NN, la de SVM y la de Árbol de Decisiones para saber, por medio de sus matrices de confusión, donde los ratios de exactitud y de error de cada regresión fueron calculados, cuál de las 4 fue la que más acertó para cada escenario.

PALABRAS CLAVES:

Condición; falla; atoro; error; exactitud; colector.

ABSTRACT

MANAGEMENT PROPOSAL FOR THE DETECTION OF FAILURES IN RESIDUAL WATER COLLECTORS OF THE DISTRICT OF SAN JUAN DE LURIGANCHO

In this thesis, three failure scenarios were evaluated in wastewater collectors of the San Juan de Lurigancho district. The first analysis consisted in evaluating the sewer's condition according to the diameter, the slope, the time (in months), the number of failures and the depth, and the second analysis consisted in evaluating if a collector suffers a stuck according to the number of activities using like variables to the diameter, the slope, the speed, the tie, the flow, the angle, the wet perimeter, the area, time (in months) and the hydraulic diameter. The analysis was performed using the method known as Machine Learning through the clustering method, what served to do the predictions in both scenarios through k-means clustering method then apply the 4 regressions which are the logistic, the K-NN, the SVM and the Decision Tree to know, thought of their confusion matrix, where the accuracy ratios and the error ratios were calculated, which of the four was the most correct for each scenario.

KEYWORDS:

Condition; failure; stuck; error; accuracy; collector.

INTRODUCCIÓN

Los servicios básicos, son obras de infraestructura de vital importancia en una ciudad, barrio o centro poblado, cuya función es brindar salubridad y una buena calidad de vida a la población. Dentro de estos servicios se encuentra el agua potable y el alcantarillado; sin embargo, en el Perú este servicio es cubierto solo en un 94% en las áreas urbanas y 71.9% en áreas rurales [1,2].

En el caso de Lima Metropolitana, la empresa que brinda el servicio de agua potable y alcantarillado en el sector urbano es la empresa estatal Peruana del Servicio de Agua Potable y Alcantarillado de Lima - SEDAPAL S.A., creada en 1981 por el Decreto Legislativo N°150 de 1981 [3].

En el año 2018, dicha empresa registró una producción total de agua potable de 729'326 m³ en Lima Metropolitana como lo señala la Tabla 1 del Anexo 1, mientras que el porcentaje de aguas residuales tratadas para dicho año fue de 92.20% según la Tabla 2 del Anexo 1.

Desafortunadamente SEDAPAL registró muchas fallas, tal como se aprecia en los datos del 2018 de la Tabla 1 del Anexo 3.

Por este motivo, el presente trabajo desarrolla un modelo predictivo para la identificación de futuras fallas en el sistema de alcantarillado de agua servidas en el distrito de SJL, para la prevención de fallas en un sector crítico de un tramo de ducto que transporta aguas servidas.

Para ello, en hará uso del método de data mining o minería de datos, con la finalidad de desarrollar una predicción con tal de ver si hay probabilidad de falla de todo los que se va a mencionar más adelante.

Alcance

El estudio está focalizado en los tramos troncales de SJL que corresponden a los colectores secundarios.

Para el análisis se consideró las características de los ductos tales como diámetros, tiempo, profundidad, material, pendiente, tirante, caudal y la velocidad del agua residual por brindada por Sedapal. Esta data fue procesada utilizando el software Python en Google Colab por el método de clusterización aplicando los métodos de las 4 regresiones que son la regresión Logística, K-NN, SVM y Árbol de Decisiones, con el cual se trabajó el modelo.

El presente estudio se limita a hacer un análisis técnico de esa data, el cual no incluye un análisis de costos, este método se llama Machine Learning.

Antecedentes

Como parte de la revisión de la literatura, a nivel internacional se han encontrado investigaciones enfocadas la detección de fallas en los sistemas de alcantarillado las cuales se detallan a continuación:

- En 2014, en la Universidad Autónoma de México, se hizo un análisis del sistema de alcantarillado, con la finalidad de prevenir futuros problemas, entre ellos fugas [4].
- En 2016, en la Universidad Católica de Colombia, se hizo un estudio y diagnóstico de la red de alcantarillado sanitario y pluvial para el proceso de un sector del centro de Bogotá. En este estudio se usó el programa EPA SWMM que permite el cálculo requerido para modelar un sistema de alcantarillado y evaluar su estado [5].
- En 2011, en la Universidad Técnica de Ambato de Ecuador, se hizo una investigación sobre un modelo de gestión y su incidencia en la provisión de servicios de agua potable y alcantarillado en la municipalidad de Tena, con la finalidad de desarrollar un plan de mantenimiento permanente para los dos servicios mencionados [6].

- En 2013, se hizo una evaluación del sistema de agotamiento sanitario de la Región Administrativa de Sobradhino cuya finalidad fue evaluar y ponderar la eficiencia de la aplicabilidad de conocimientos y el desarrollo de nuevos métodos y tecnologías para la utilización correcta del sistema de alcantarillado de la región de Sobradhino [7].
- En 2012, en el Instituto Politécnico y Universidad Estatal de Virginia, se hizo uso del método de data mining para las tuberías a gran escala concluyendo que se necesitan prácticas en las evaluaciones de la condición de sus programas de gestión de activos [8].

Por otro lado, en el Perú, se tiene la siguiente investigación:

- En 2018, en la Universidad Cesar Vallejo, se hizo una evaluación del sistema de alcantarillado con la aplicación del software SewerCad en la prolongación los Libertadores en Huaraz. En esa investigación se concluyó que por medio del programa mencionado se puede minimizar costos y reducir el movimiento de tierras, el cual resulta tener mayor incidencia en el costo de un sistema de alcantarillado combinado [9].

A diferencia de la investigaciones encontradas, la presente investigación, usará la minería de datos o data mining con el fin de hacer el análisis de condición que permita detectar las fallas.

Justificación y motivación

El de esta investigación es desarrollar un modelo predictivo por medio del método predictivo de machine learning que permita anticipar las fallas en los colectores con la finalidad de prevenir los incidentes, debido a que un aniego que contiene aguas residuales es peligroso para el ser humano por la presencia de contaminantes, pudiendo traer como consecuencia enfermedades como infecciones de piel, infecciones de tracto genitourinario, leptospirosis, cefalea, crisis de migraña, síndrome de ansiedad y depresión, terror nocturno

en los niños, heridas infectadas, tétanos, deshidratación, enfermedades respiratorias agudas, conjuntivitis, dengue y cólera sumado el corte y la ausencia del servicio de agua potable.

Además de los problemas a la salud, la población se ve afectada por la incapacidad de acudir a sus centros laborales y ausencia escolar.

Asimismo, los perjuicios ocasionados a las personas generarían a la empresa SEDAPAL, costos significativos por retribución de daños en las viviendas y enseres, disponibilidad del personal para la reparación y mantenimiento, personal encargado para el monitoreo del restablecimiento de agua en los colegios, entre otros [10].

La técnica para el análisis es utilizando Machine Learning que consiste en el análisis de base de datos históricos que permita descifrar comportamientos en el futuro. Para ello se han definido algunas variables como: La clasificación de la condición del colector y su predicción a través de las 4 regresiones como variables dependientes y como variables independientes: tiempo del colector, diámetro del colector, pendiente del colector, material del colector, profundidad del colector, tirante del agua residual, velocidad del agua residual, y el caudal del agua residual.

Mitigar este problema a través de un mejor enfoque que beneficiará a las empresas que las realizan los estudios, los resultados del estudio permitirán facilitar la detección de fallas antes de generar un problema a la sociedad.

El estudio propondrá un modelo predictivo para los colectores secundarios Independencia que permita tomar acciones antes que ocurra averías de diferente índoles, beneficiando a SEDAPAL a evitar inundaciones en el colector de dicha avenida como consecuencia de aplicar acciones correctas.

Objetivo general

Desarrollar un modelo predictivo para la detección de fallas en colectores de aguas residuales del distrito de SJL orientada al mantenimiento.

Objetivos específicos

- Categorizar los distintos tipos de fallas que ocurren en los ductos de aguas servidas del distrito de SJL.
- Analizar el impacto técnico de las fallas encontradas.
- Aplicar metodologías de base de datos para evaluar la probabilidad de una falla.

CAPÍTULO I

MARCO TEÓRICO

El presente capítulo compila toda la información necesaria para poder identificar oportunidades de mejora de los métodos de mantenimiento de los colectores del distrito de SJL, la cual está relacionada infraestructura, topologías, materiales, etc. del sistema de alcantarillado.

1.1 Sistema de distribución de Aguas servidas

Se define como aguas servidas a aquellos líquidos que son el resultado del uso del agua el área doméstica o industrial. Existen siete tipos de aguas residuales:

- a) Industriales: Efluentes de las operaciones de las industrias.
- b) Domésticas: El origen de las aguas residuales son producto de las aguas utilizadas en las casas, centros de estudio, etc.
- c) Municipales: Son las que circulan por el tramo de una ciudad o población.
- d) Agua negra: Son las que contienen los desechos de las actividades de los baños, como orina y heces.
- e) Agua amarilla: Es cuando la orina es transportada con o sin la presencia de agua.
- f) Agua café: Es el tipo de agua residual que contiene pequeñas cantidades de heces y orina.
- g) Agua gris: Su origen es de las duchas, lavamanos y lavadoras [11].

El sistema de distribución de aguas servidas, también llamado sistema de alcantarillado, son redes de instalaciones de tuberías destinadas a la recolección, evacuación y disposición final [12].

Por otro lado, los alcantarillados están clasificados de acuerdo al tipo de agua residual que transportan:

- a) Alcantarillado Sanitario: Encargados de la evacuación rápida y segura de las aguas servidas de los casas, lugares comerciales o plantas donde se realizan tratamiento.
- b) Alcantarillado Pluvial: Se encarga de la captación y conducción el agua de la lluvia hacia su evacuación final.
- c) Alcantarillado Combinado: Este sistema conduce por completo las dos aguas mencionadas anteriormente, pero presentan dificultades después de su tratamiento y pueden ocasionar problemas al medio ambiente si el agua del alcantarillado combinado hace contacto con cauces naturales.
- d) Alcantarillado Semi-combinado: Este sistema conduce el 100% de las aguas residuales producidos en uno o varios sectores, en porcentajes menores al total de las aguas pluviales captadas en los diferentes sectores y con conducciones que brinden una mejoría y complemento al sistema pluvial con el fin de evitar inundaciones en áreas pobladas.

1.1.1 Componentes

Los principales componentes de un sistema de alcantarillado son:

- a) Colectores principales: Son los que presentan mayor diámetro, ocasionalmente se ubican en los lugares donde el nivel del poblado es bajo. Estos colectores transportan las aguas servidas hasta un destino establecido.
- b) Colectores secundarios: Estas tuberías están conectadas a los colectores principales, son las que recolectan las aguas servidas de los domicilios conduciéndolas al colector principal.
- c) Cajas de inspección: Son cámaras que permite el ingreso a los colectores facilitando el mantenimiento de las mismas [13].

1.1.2 Topología

La topología se aplica en todas las áreas de la industria, pues existen diversas formas de interconectar varios sistemas. Un sistema está integrado de varios componentes o dispositivos; a medida que el sistema crece, es cuando se debe considerar la topología. Así, el sistema de alcantarillado tiene sus propias topologías:

- Perpendicular

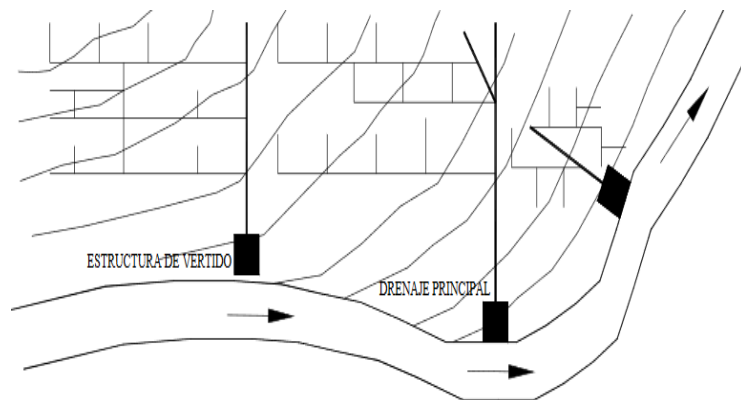


Figura 1.1 Topología perpendicular [14]

- Radial

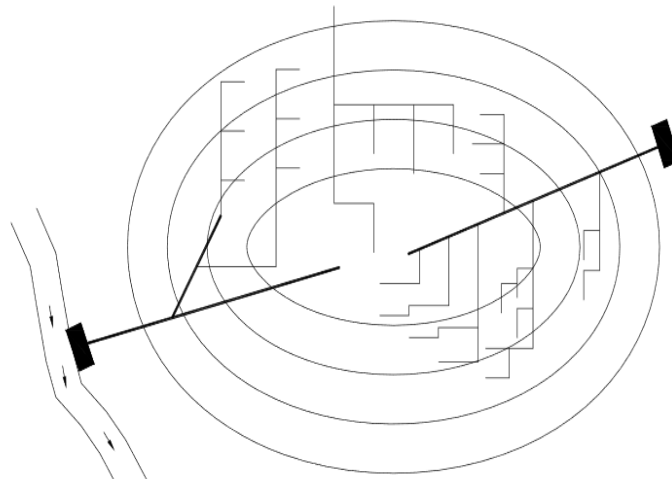


Figura 1.2 Topología radial [14]

- Interceptores

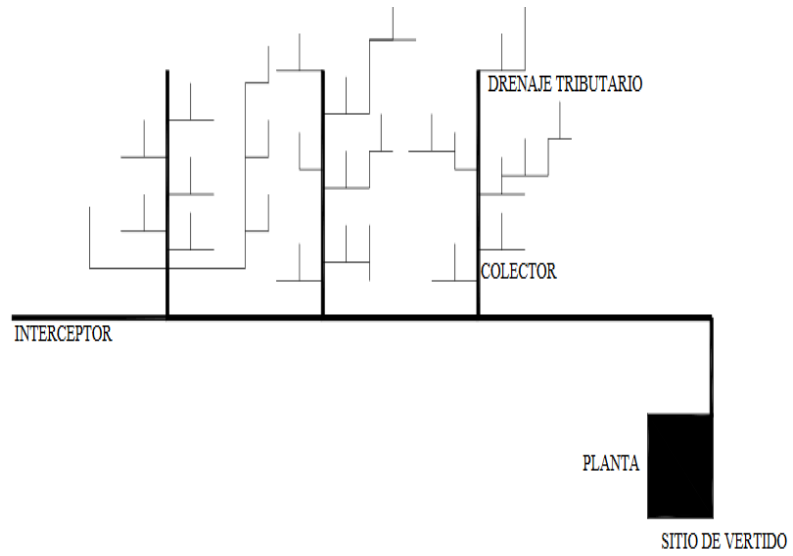


Figura 1.3 Topología de Interceptores [14]

- Abanico [14]

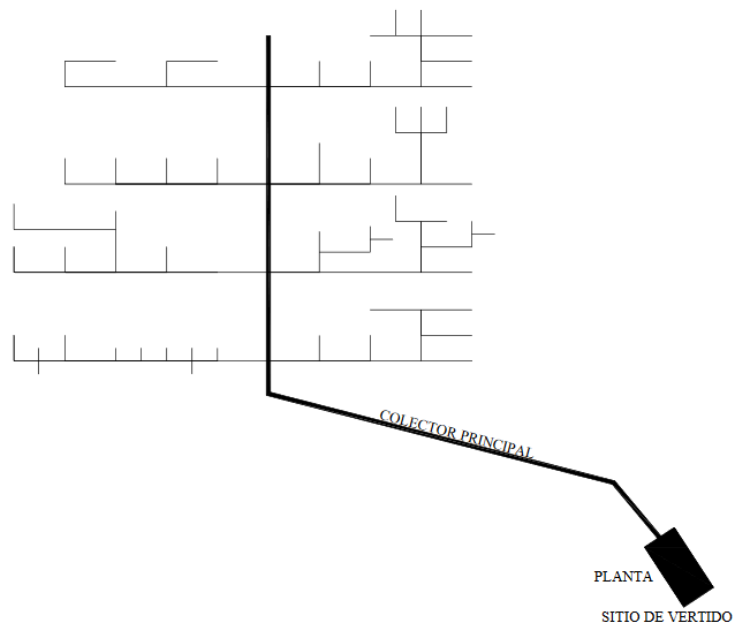


Figura 1.4 Topología de Abanico [14]

En la figura 1.5 se muestra la topología del sistema de alcantarillado de SJL, así como el colector que pasa en el distrito mencionado, que es el Colector Canto Grande. Además, se puede ver que la topología que aplica el sistema de alcantarillado de SJL es de tipo radial.

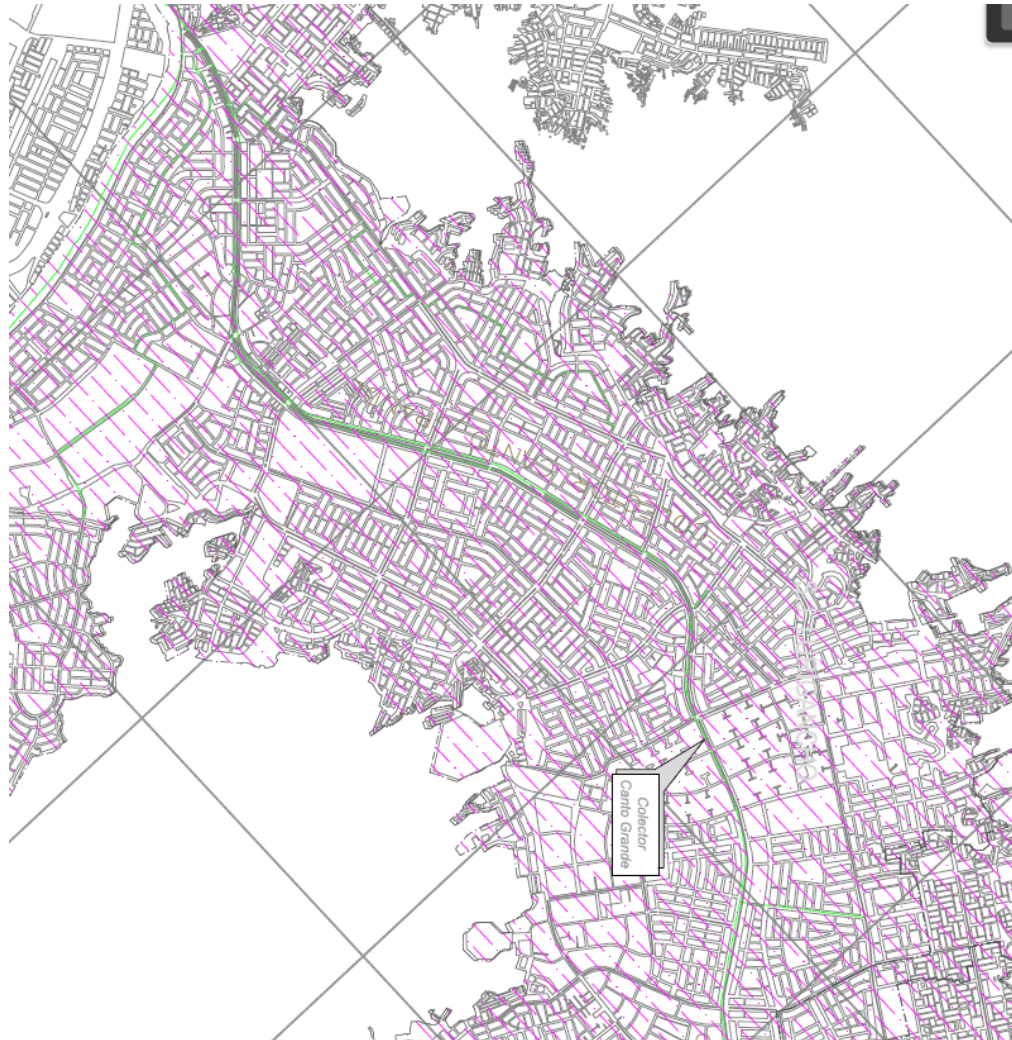


Figura 1.5 Plano del sistema de alcantarillado de SJL [15]

1.1.3 Protocolos de instalación y pruebas

Para el desarrollo el modelo que ayudará en la detección de fallas, es necesario conocer los protocolos que se aplican. Esta información servirá para conocer la ubicación del tramo de colector que va a ser analizado.

1.1.3.1 Protocolos de instalación

Para la instalación o cambio de una tubería se ejecuta las siguientes acciones:

- a) Trazo: Para llevar a cabo un trabajo, se hace un trazado de la zona donde se va a realizar dicha tarea, usando yeso o cal.
- b) Corte y Rotura de Pavimento: Para realizar el mantenimiento en una zona donde existe pavimento, se realiza un proceso de corte con equipo de disco diamantado. El corte debe ser profundo que corte casi todo el espesor del pavimento o con una profundidad que no sea menor a los 10 cm. Este proceso de rotura puede ser hecho de manera manual o a través del uso de martillos neumáticos.
- c) Proceso de excavación de la zanja: Puede ser de dos modos: manual o mecanizado. Para el modo manual se utiliza lampas y picos; para la mecanizada, la retroexcavadora.

Para que el trabajo se haga de manera más segura, las dimensiones de la zanja deben ser adecuadas, con la utilización correcta de las herramientas de tal forma que permita el correcto uso de los materiales de acuerdo a las especificaciones técnicas y sugerencias de los fabricantes.

El material que ha sido extraído debe colocarse a un lugar distante del área de trabajo (mínimo a un metro del borde de la zanja) de manera que no perjudique el trabajo que se realiza en la zanja. El material extraído que esté mojado o sea de mala calidad, debe ser desechado, ya que no tiene utilidad en el rellenado de la misma zanja de donde se extrajo.

- d) Proceso de entibado de zanjas: Este procedimiento es utilizado cuando el suelo es deleznable y susceptible al desmoronamiento con el fin de proteger a los trabajadores y a las tuberías.

Los elementos que se usan para este proceso tienen sus propios nombres de acuerdo a la posición de la zanja tales como:

- Estacas: Colocadas de forma vertical. El largo usado para la colocación de la estaca tiene un nombre propio, llamado ficha, pero en el caso de que la tierra empuje de manera directa, se llamaría tablestacado.
 - Vigas: También denominado soleras, son colocadas de manera longitudinal y están en paralelo con el eje de la zanja.
 - Puntal: Son capaces de transmitir la fuerza que ejerce la tierra al empujarla desde un lado de la zanja hacia el otro lado. Son colocados de manera transversal y es de costumbre empalar con puntales rollizos.
- d) Proceso de tablestacado: El tablestacado consiste en la colocación de tablas apoyadas en las dos paredes de la zanja y están sostenidas entre sí con maderas o troncos que tengan una longitud igual al ancho de la zanja, evitando estorbar proceso de ejecución del servicio. Para elegir el tipo de tablestacado que va a ser utilizado, es necesario tener en cuenta los siguientes parámetros:
- Tipología del suelo
 - Profundidad y ancho de la zanja
 - Tiempo en que la zanja estará abierta
 - Condición de los bordes laterales
 - Riesgo de la infiltración de agua de las lluvias, etc.
- e) Drenaje de Zanja: Es necesario que la zanja esté libre de la presencia de agua. Para ello se está haciendo el uso de bombas a diésel o gasolina, no es sugerible el uso de una bomba eléctrica.

En caso de que no disponer de una bomba, se puede usar un balde de 20 litros.

Las aguas extraídas deben ser eliminadas a través de un colector, nunca deben ser vertidas en la calle.

- f) Ejecución efectiva de los servicios: La instalación de las tuberías, construcción y rehabilitación de buzones, así como las reparaciones de las conexiones domiciliarias son descritas y establecidas.

También se debe aclarar la intervención de otras variables durante la ejecución, como: el tipo de material, diámetro, longitud, profundidad y tipo de ocurrencia. En la Tabla 1 del Anexo 2 se puede apreciar un ejemplo.

- g) Relleno de las Zanjas: Una vez concluida la reparación o instalación después que se haya verificado que no haya presencia de fugas en las uniones o empalmes, se debe realizar este trabajo.

Si el material extraído no es apropiado para el proceso de rellenado, se deberá utilizar un material que sea adecuado y proveniente de un área que lo tenga disponible.

El rellenado se lleva a cabo con un material escogido para el compactado por capas a través de procesos manuales o mecánicos.

- h) Proceso de eliminación del desmonte y material residual: Después del rellenado, el material excedente de la excavación realizada, deberá ser retirado del lugar donde se llevó a cabo el servicio.

Los desmontes, junto con el material excedente, son eliminados conjuntamente en el caso de una reparación, sustitución de tuberías, sustitución de conexiones domiciliarias y rehabilitación de buzones.

- i) Restablecimiento del nuevo pavimento: El pavimento que pasó por el proceso de corte y rotura, debe ser dejado en las mismas o mejores condiciones que tenía el pavimento previamente.

Antes del restablecimiento del pavimento, es necesario que la base sea preparada colocando una capa de 20 cm de un material afirmado y con la compactación adecuada [16] [17].

1.1.3.2 Protocolos de pruebas

El objetivo de estos estudios es asegurarse de que todas las partes de agua potable hayan sido instaladas de manera correcta con la finalidad de que estén listas para brindar los servicios. Los resultados están dirigidos y son verificados por la Supervisión con asistencia del constructor, este último es responsable de brindar el personal, material, aparatos de prueba, aparatos de medición y cualquier otro elemento que sea necesario para realizar estos estudios.

Para el sistema de alcantarillado se aplica la prueba de zanja abierta y de acuerdo con las condiciones que pueden presentarse en la obra.

Las pruebas que se ejecutan son:

- a) Pruebas de nivelación y alineamiento: Se efectúan con el uso de instrumentos topográficos, de nivel óptico, pudiendo hacer uso del teodolito cuando los tramos presenten muchos cambios de estación.

Para una pendiente superior a 10 0/00, el margen de error permitido deberá ser mayor que la suma algebraica ± 10 mm. La medición es realizada entre dos o más puntos. Por otro lado, Para una pendiente menor a 10 0/00, el margen de error permitido no debe ser mayor que la suma algebraica de \pm la pendiente. En ambos casos, se realiza entre dos o más puntos.

En el caso de las líneas con tubería flexible, la prueba puede ser realizada por el método fotográfico, con CCTV o a través de espejos colocados a 45° , lo que debería permitir la visualización del diámetro de la tubería al hacer la observación de buzones consecutivos.

- b) Prueba hidráulica: En un sistema de alcantarillado no se puede realizar este tipo de prueba si es que no ha cumplido de manera satisfactoria la prueba a zanja abierta. Las pruebas hidráulicas son de dos tipos:
 - Prueba de Filtración, cuando la tubería ya fue instalada en terrenos secos sin presencia de agua freática.

- Prueba de Infiltración, que es usada en los terrenos con presencia de agua freática.
- b) Prueba de filtración: Prueba que examina que no haya presencia de fugas en las uniones o en el cuerpo de la tubería y que el buzón de prueba presente lecturas correctas.

En esta prueba se usa agua limpia para el llenado del buzón hasta su altura total y conveniente tapado en el buzón de aguas abajo. El agua permanecerá al menos 24 horas en el tramo para que la prueba pueda ser realizada.

La prueba de zanja abierta, no debe presentar ningún relleno en el tramo, en otras palabras, debe estar libre, con las uniones totalmente descubiertas. Así mismo, hasta que la prueba no haya terminado, no se deben llevar a cabo los anclajes de buzones y/o de las conexiones domiciliarias.

La duración de la prueba no debe ser menor a 10 minutos, en los tramos que poseen tubos cuyo material no absorba agua y no se admita pérdidas en el tramo probado.

También puede realizarse la prueba de filtración de manera práctica midiendo la altura que baja el agua en el buzón en un tiempo determinado.

En las pruebas de relleno compacto, en donde también están incluidas las pruebas de caja de registro, se puede realizar el mismo procedimiento que las pruebas a zanja abierta [18].

1.1.4 Materiales en tuberías

Existen tuberías hechas de distintos tipos de materiales que se emplean para el transporte de aguas servidas. Esta información sirve para tener una idea de qué tipo de material usa la tubería que transporta las aguas servidas, ya que el tramo que va a ser analizado tiene su respectivo material y diseño.

Los materiales que se emplean en las tuberías tienen diferentes cualidades como duración, resistencia estructural del material, resistencia mecánica, conductividad, características de los suelos, características del agua, costos de los materiales, de fácil manipulación, el colocado y su instalación, de diseño flexible y el mantenimiento y la reparación sencilla.

Los que son usados para fabricar las tuberías son [19]:

- a) Acero
- b) Concreto simple (CS) y concreto reforzado (CR)
- c) Concreto reforzado con revestimiento interior (CRRI)
- d) Poliéster reforzado con fibra de vidrio (PRFV)
- e) Policloruro de vinilo (PVC) (pared sólida y estructurada)
- f) Fibrocemento (FC)
- g) Polietileno de alta densidad (PEAD) (pared sólida corrugada y estructurada)

1.2 Servicio de alcantarillado en San Juan de Lurigancho

Como se mencionó en la parte de la topología, el colector principal de SJL es el colector Canto Grande. El distrito de SJL, según la INEI, es el distrito más poblado de todo Lima Metropolitana. En la siguiente figura se puede apreciar el nivel de población que tiene el distrito [20].

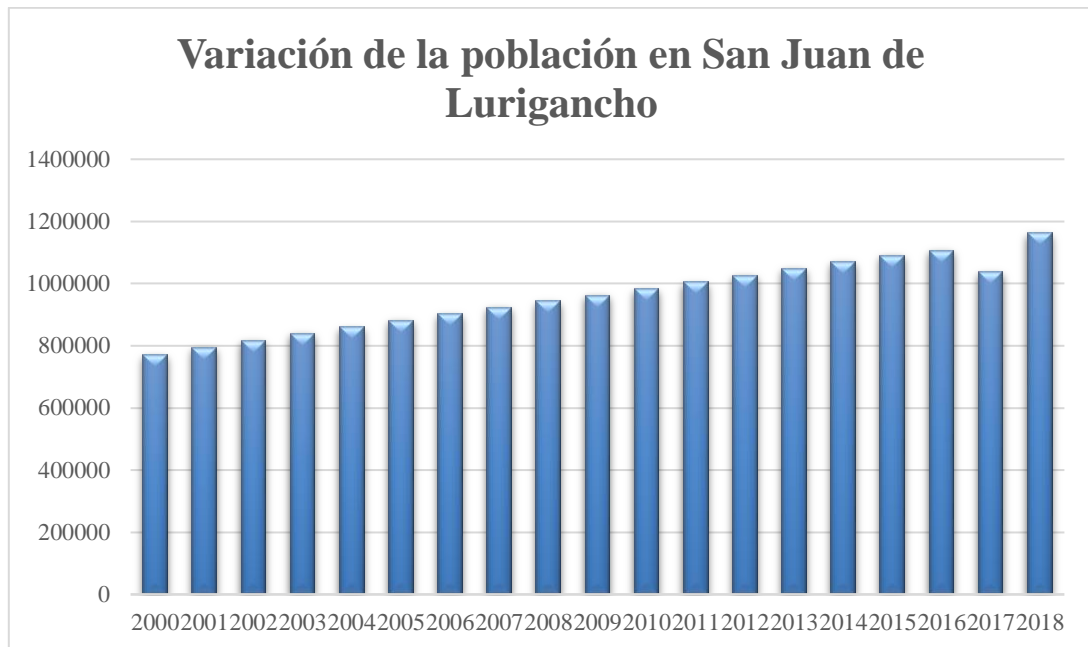


Figura 1.6 Población de San Juan de Lurigancho [20] [21]

Según el reporte de la INEI del 2017, tan solo el 6.9% de la población de SJL cuenta con servicio del agua potable [22].

El trazo del colector Canto Grande recorre las avenidas de Wiese, Próceres de la Independencia, Santa Rosa, Marañón y vía Evitamiento, hasta desembarcar a la altura de la Plaza Acho en el Colector N°6, recolectando buena parte de los desagües de SJL. Los diámetros de este colector son 600, 700, 800, 900, 1000, 1300 y 1200 mm. En su trayecto, este colector cuenta con los siguientes colectores:

- **Emisor Wiese:** Recolecta los desagües de la parte alta del distrito de SJL, de los A.H., Montenegro, Cruz de Motupe, Isabel Flores de Olivia, José Carlos Mariátegui y Ciudad Satélite Canto Grande, entre las principales. Presenta diámetros de 350, 400, 450, 525 y 600 mm; el material que se usa para el tramo es el concreto y el PVC y su trazo es a lo largo de la Av. Wiese, desde A.H. Cruz de Motupe hasta la Av. Bayóvar, descargando en el colector principal ubicado en esta última avenida.

- Colector Bayóvar: Recolecta los desagües de la ciudad satélite de Canto Grande, A.H. 27 de marzo y Urb. Andrés Avelino Cáceres, entre los principales. Su tramo recorre la avenida Bayóvar y posee un diámetro de 400 mm hasta descargar en el colector principal ubicado en esta última avenida.
- Colector Las Flores: Su descarga se efectúa en la intersección Madre de Dios con la Av. Próceres de la Independencia con una diámetro 450 mm y recolecta los desagües de las urbanizaciones Las Flores, Ricardo Palma y Chacarilla de Otero [23].

1.3 Fallas en sistemas de distribución

Para el desarrollo del modelo es necesario conocer cuáles son los tipos de fallas por las que puede pasar una tubería. A continuación, se detallan las causas de daños del sistema de alcantarillado.

1.3.1 Causas de daños en el sistema de alcantarillado

Entre las causas de fallas o daños del sistema de alcantarillado se encuentran:

- a) Material de construcción: Cuando la tubería es construida con materiales de mayor tendencia a la corrosión son potencialmente susceptibles a un mayor deterioro o derrumbamiento. Un ejemplo de un material susceptible a la corrosión es el concreto no reforzado.
- b) Pendiente mínima: Cuando las tuberías están ubicadas en áreas de reducida inclinación, hay una alta probabilidad de que puedan ocurrir inundaciones.
- c) Acumulación de Material: La acumulación de material puede ocasionar atoros. Esto puede ocurrir en tuberías con poca pendiente y con velocidades de caudales muy bajos.
- d) Obstrucciones: Producidos por las acumulaciones de diferente tipo de materiales, ya sea a nivel de tuberías o buzones de retención de sólidos. Las obstrucciones aumentan en tuberías donde la velocidad del caudal es muy baja.

- e) Grasas: Las obstrucciones también se dan donde en los mercados y restaurantes. Cuando las grasas están presentes en el alcantarillado, estas se endurecen y con el tiempo forman tacos de sebo que obstruyen las tuberías. La incidencia de la presencia de grasa es mayor en tramos donde la pendiente es baja.
- f) Trapos y plásticos: Estos materiales se encuentran a menudo cuando una tubería es obstruida. En las zonas donde hacen mal uso del sistema de alcantarillado es mayor su incidencia.
- g) Raíces: Este tipo de daño ocurre mayormente en las zonas verdes con árboles, lo que hace que las raíces obstruyan el alcantarillado. Las raíces de diferentes tipos de plantas tienden a penetrar en las uniones o en las roturas de las tuberías ocasionando obstrucciones completas.
- h) Roturas: Este tipo de eventos y fallas que se presentan en el sistema de alcantarillado se pueden deber a:
- Base y soporte inapropiado del tubo: La falla ocurre debido a que una tubería es colocada en una zanja con un fondo no uniforme y sin una buena cama de apoyo.
 - Cargas vivas: Las tuberías que tienen un inapropiado recubrimiento presentan gran probabilidad de colapsar debido a la sobrecarga a la que está sometida, sobre todo si está en una zona donde el tráfico es de vehículos pesados.
 - Movimiento de tubo: Estas fallas son ocasionadas por los sismos o terremotos.
 - Otras causas: Conexiones domiciliarias no autorizadas, Conexiones con desechos industriales y condiciones deficientes en los procesos de construcción [24].

1.4 Herramientas de diagnóstico para la detección de fallas

Para las inspecciones en los colectores es necesario conocer qué herramientas son empleadas a la hora de poner en práctica, ya que para la presente tesis es necesario conocer qué herramientas se usan en la inspección de los colectores.

1.4.1 Inspección por medio de circuito cerrado de televisión

Existen equipos que son utilizados para la detección de fallas o fugas en una tubería del sistema de alcantarillado; esta inspección se realiza por medio de equipos CCTV.

Los equipos CCTV cuentan con dos sistemas básicos para realizar las inspecciones. Primero se tiene a los sistemas estacionarios con cámaras de muy buena capacidad de acercamiento (zoom). Estos son colocados en los pozos de inspección, ya que se puede tomar imágenes desde allí; es difícil la cuantización de las fallas que están alejadas de la cámara, ya que este tipo de sistema solo provee información cualitativa del estado de las fallas dentro de una tubería. Se puede emplear un sistema inspección preliminar con la finalidad de determinar qué zonas necesitan de información adicional que se pueda obtener con un sistema móvil.

El segundo tipo es el sistema móvil de inspección de CCTV, el cual está compuesto por una cámara acondicionada con un sistema de propulsión que puede estar montada sobre un transportador. Estas cámaras poseen mecanismos de rotación y zoom, son muy útiles para tomar de cerca los defectos encontrados. Pueden ser transportados por medio robots, llantas adheridas u orugas que poseen mecanismos que permiten el movimiento en distintas direcciones y el control de la velocidad de avance del sistema. Este sistema es ventajoso para los operadores que quieren realizar las inspecciones a su propio ritmo dependiendo la experiencia que poseen.



Figura 1.7 Cámaras montadas sobre transportadores [25]

1.4.1.1 Equipos necesarios para la inspección

Para el sistema CCTV, las cámaras de televisión y los transportadores necesitan otros equipos especiales para completar el sistema y la adecuada recopilación de información, entre los que se encuentran:

- a) Carretes de Cableado: Estos transportan cables para brindar la energía necesaria al sistema de cámara para su movilización y funcionamiento. Los cables también sirven para transmitir la información recopilada por la cámara a la unidad o al centro de mando. Este sistema puede estar limitado por la longitud del tramo donde se realiza la inspección.
- b) Centro de mando: Es en donde el sistema es operado y se recopila la información que es registrada por la cámara.
- c) Sistemas de iluminación: La iluminación es el sistema más importante, ya que el sistema debe tener la iluminación suficiente para que todos los detalles puedan ser captados con una buena calidad de imagen y nítida.

1.4.2 Otros métodos para la inspección de la condición interna de una tubería

1.4.2.1 Sistema de escaneo por láser

Este sistema puede medir los perfiles de las tuberías con una precisión mayor a 0.1 cm, tiene procesadores computacionales, lo que permite analizar la información conforme la inspección sea realizada.

Es útil para determinar el diámetro real de la tubería, la profundidad del flujo, el espesor de capas de sedimentos, las deformaciones causadas por la erosión, el desplazamiento y separación de juntas, etc.

En este sistema, los sistemas CCTV se usan en conjunto y es una herramienta bastante precisa y práctica, ya que permite cuantificar las fallas presentes en el colector.

1.4.2.2 Sistema de escaneo por Sonar

En este sistema se emplea el ultrasonido y permite establecer el perfil de la tubería. Su principal ventaja, frente a los métodos visuales anteriormente mencionados, es que recopila información debajo de la superficie del flujo, lo que permite su utilidad en tuberías por donde el caudal no puede ser detenido.

El sistema sirve para la identificación de problemas como sedimentación, capas de grasa, niveles de bloqueo. Mediante un software especializado los problemas pueden ser cuantificados, incluso determinando el espesor de la tubería con el uso de relaciones geométricas con el diámetro. Algunos programas poseen interfaces gráficas que permiten calcular el cambio en el perfil de la viga antes y después de su deformación [25].

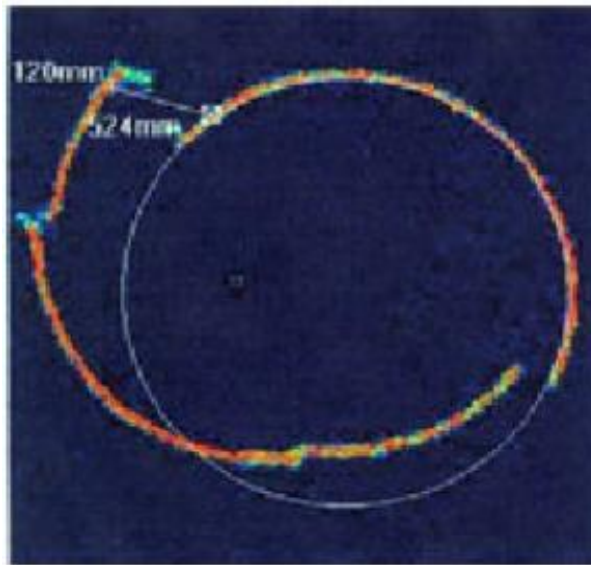


Figura 1. 8 Comparación de diámetro teórico con diámetro original [25]

1.5 Herramientas de diagnóstico por base de datos

El Data Mining (DM) es la obtención información objetiva en amplias bases de datos. Su nombre procede de la semejanza existente entre buscar información de valor para un determinado usuario y el proceso de minado en una montaña para encontrar una cantera con metales preciosos, ya que estos procesos requieren de la evaluación de una gran de cantidad de material e investigar de forma inteligente hasta terminar la búsqueda.

Es un grupo de herramientas que ayuda a enfocarse los objetivos que desean lograr sobre la información de mayor importancia de una fuente de datos.

Las herramientas que brinda el DM sirven para responder las interrogantes que toman demasiado tiempo en responder.

El procedimiento consiste en la repetición de datos con el fin de descubrir patrones y tendencias, mediante ordenes automatizados, semiautomatizadas o manuales y que no serían descubiertos por los métodos tradicionales de estadística para una investigación.

Esta herramienta permite realizar una exploración de una base de datos con la finalidad de hallar patrones ocultos, permitiendo la predicción de futuras tendencias y comportamientos de información descubierta recientemente.

Cuando un nuevo patrón es descubierto, este debe alguna ventaja.

Existen extensas fuentes de datos almacenados en fuentes de información, que van creciendo y esa información podría ser usada en algún momento. Sin embargo, la diferencia entre generación de datos y entendimiento de datos crece progresivamente; en otras palabras, así mientras la cantidad de datos esté en aumento, menos será la cantidad de personas que lleguen a entender de manera satisfactoria. Por el contrario, en algunas ocasiones la adquisición de datos no es tan sencilla

Por otro lado, la información que está encubierta en los datos es beneficioso y por lo general no se encuentra de manera evidente a fin de poder obtener algún provecho para la finalidad requerida.

En algunas situaciones, hay información que no puede ser analizada a través de los métodos estadísticos ya existentes, ya que pueden encontrarse valores perdidos o bien, los datos pueden estar de manera cualitativa o no cuantitativa.

Ya en la actualidad, es necesario usar las herramientas del DM para el desarrollo del almacenamiento de datos (Data Warehouse) a gran escala [26].

1.5.1 Ventajas y Desventajas

a) Ventajas:

- Se descubre información no esperada. Se descubren relaciones que son difíciles de obtener por otros métodos que sirven de aporte para la empresa.
- Los métodos empleados son comparados con técnicas estadísticas antes de que sean usados asegurando una gran fiabilidad.

- Los modelos se pueden construir rápidamente. Permite la generación de modelos en tan solo unos minutos. Los modelos son probados y solo el mejor es entregado al usuario.

b) Desventajas:

- Dificultad en la recopilación de datos. Dependiendo del tipo de dato, puede conllevar un gran trabajo obtenerlos y procesarlos, o tener un elevado coste de agrupación.
- Se requiere una gran inversión de tiempo en el procesamiento de los datos obtenidos [27].

1.6 Método de Análisis de falla

1.6.1 Machine Learning

El Data Mining tiene muchos métodos, uno de ellos es el Machine Learning (aprendizaje automático). Este método se centra en la visión de la computadora, audición, el procesamiento de los lenguajes naturales, procesamiento de imágenes y reconocimiento de patrones, computación cognitiva, representación del conocimiento, etc.

Estas tendencias de investigación tienen como objetivo proporcionar a las máquinas la capacidad de recopilar datos a través de sentidos similares a los sentidos humanos y luego procesarlos mediante el uso de herramientas de inteligencia computacional y métodos de aprendizaje automático para realizar predicciones y tomar decisiones al mismo nivel que las personas [28].

Los sistemas del aprendizaje automático normalmente se clasifican por sus estrategias de aprendizaje subyacente que a menudo se identifican por la cantidad de inferencia que el programa de computadora es capaz de realizar:

- a) Aprendizaje de roles: Llamado Role Learning en inglés, cuya función es describir la estrategia que utilizan los programas tradicionales. No realizan ningún tipo de inferencia y todo su conocimiento tiene que ser implementado directamente por el programador, ya que la aplicación no puede sacar conclusiones o transformaciones de la información dada.

- b) Aprendiendo de la instrucción: Llamado Learning from Instruction en inglés, abarca todos los programas de computadora que pueden transformar información del idioma de entrada dado un idioma interno. Aunque el programador todavía proporciona el conocimiento sobre cómo realizar esta transformación de manera efectiva, esto requiere pequeñas formas de inferencia desde el lado del programa de computadora. Por lo tanto, se define un nivel separado de sistema de aprendizaje en comparación con el aprendizaje de memoria.
- c) Aprendizaje por analogía: Llamado Learning by Analogy, a diferencia del método anterior, este trata de desarrollar nuevas habilidades que son casi similares a las habilidades existentes y, por lo tanto, fáciles de adoptar, realizando transformaciones en la información conocida. Este sistema requiere la capacidad de crear mutaciones y combinaciones de un conjunto dinámico de conocimientos. Crear nuevas funcionalidades, que eran desconocidas para el programa de computadora original y, por lo tanto, requieren inferencia.
- d) Aprendiendo de ejemplos: Llamado Learning from examples en inglés, en la actualidad es una de las estrategia de aprendizaje más utilizadas, ya que proporciona la mayor flexibilidad y permite que los programas informáticos desarrollen habilidades completamente desconocidas o encuentren estructuras y patrones desconocidos en los datos. Aprender de los ejemplos es una técnica que a menudo se usa en tareas de clasificación y minería de datos para predecir la etiqueta de clase de las nuevas entradas de datos basadas en un conjunto dinámico de ejemplos conocidos.

1.6.2 Clusterización, K-mean y las 4 regresiones

1.6.2.1 Clusterización

El estudio por medio de la clusterización concierne a una gran diversidad de métodos que pueden ser usados para tener una nueva clasificación. Los métodos clúster son métodos estadísticos de múltiples variables que empiezan con un grupo de datos que contiene

información sobre una muestra objetivo e intentar reordenarlas en grupos aproximadamente similares a los conocidos como clústeres.

A través del estudio clúster, la información es relativamente poco conocida sobre la organización de las categorías, lo cual hace que haya disparidad entre los procedimientos de múltiples variables de discriminación y asignación, donde esta información se es previamente conocida. Todo lo que se puede encontrar es una colección de observaciones, siendo la finalidad de la operación en este caso, el descubrimiento de las estructuras de las categorías en la que coinciden con las observaciones.

El objetivo de estas técnicas clusters es ordenar la observaciones en grupos tales que el grado de asociación natural sea alto entre los miembros del mismo grupo y bajo entre miembros de grupos diferentes.

Las técnicas clúster están clasificadas de diversas formas. Hay tres tipos de clasificación de estas técnicas propuestas por Kumar et al (2006), las cuales son:

- a) Jerárquicos y no jerárquicos (particionales): La diferencia entre los dos es el proceso de agrupamiento de datos, considerando si los elementos se van anidando o no.
- b) Exclusivos, Superpuestos y Difusos: Cuando los objetos son fijados en un solo grupo, se les denomina técnicas exclusivas. Cuando los objetos pueden ser fijados en un solo grupo, se les denomina técnicas superpuestas. Cuando todos los objetos pueden pertenecer a todos los grupos, asignándoles una probabilidad o peso para que sean miembros de cada grupo, se les denomina técnicas difusivas.
- c) Completo y Parcial: Cuando todos los grupos son asignados a un solo grupo, esos son los completos; en el caso de los parciales algunos objetos no serán asignados a algún grupo específico.

Ahora se va a hablar de los métodos de agrupamiento:

- a) Métodos jerárquicos: Su objetivo es la agrupación de clústeres (grupos) con el objetivo de la formación de uno nuevo o bien la separación de alguno existente con el fin de originar otros 2, de tal manera que minimice alguna función de distancia o se maximice alguna medida de similitud.

Están divididos en 2, en disociativos y aglomerativos.

- Aglomerativos: Estos empiezan el análisis con la cantidad de grupos e individuos que son el objeto del estudio. En adelante se irán formando grupos de manera ascendente, hasta el término del proceso, todos los casos están comprendidos en un mismo conglomerado.
- Disociativos: También son llamados métodos divisivos; este proceso se realiza de manera inversa al anterior. A partir del primer grupo se van formando grupos cada vez más pequeños por medio de sucesivas divisiones en grupos cada vez más pequeños. Al término del proceso se tienen tantos grupos como individuos en la muestra analizada.

Independientemente del proceso de agrupamiento, hay otros criterios para la formación de conglomerados; todos estos puntos de vista están basados en una matriz de distancias o similitudes. Por ejemplo, dentro de los métodos aglomerativos destacan:

- Algoritmo del enlace simple (vecino más cercano)
- Algoritmo del enlace completo (vecino más lejano)
- Algoritmo del promedio entre grupos
- Algoritmo del centroide
- Algoritmo de la mediana
- Algoritmo de Ward.

Dentro de los métodos disociativos, además de los métodos anteriores, los destacados y los que siguen siendo considerados válidos son los siguientes:

- El análisis de asociación
- El detector automático de interacción.

Según Pedret et al. (2003) estos métodos son apropiados principalmente para la clasificación de productos, marcas, empresas, etc. (número recluido de la data) de acuerdo a sus semejanzas sobre un conjunto de atributos o peculiaridades.

En cambio, cuando los datos están en etapa exploratoria, cuando se no se conoce o no hay una noción de la cantidad de grupos que se puedan

configurar, para encontrar el número de clústeres apropiados a partir de los datos es recomendable aplicar un método jerárquico.

- b) Métodos no jerárquicos: Son llamados métodos partitivos o de optimización, su objetivo es realizar una sola partición de los individuos en K grupos. Esto implica que el investigador debe especificar a priori los grupos que deben ser formados, siendo esta, posiblemente, la principal diferencia respecto de los métodos jerárquicos, no obstante, hay que señalar que hay diversas versiones de estos de estos procedimientos que flexibilizan un tanto el número definitivo de clústeres a obtener. La asignación de individuos a los grupos se hace por medio de algún proceso que mejore el punto de vista respecto a la selección. Otra discrepancia de estos métodos, a diferencia de los jerárquicos, se encuentra en que trabaja con la matriz de datos original y no hacen exactitud de su conversión en una matriz de similitudes o distancias.

Según Pedret et al. (2003), estos ordenamientos están agrupados en 4 métodos:

- Métodos de Reasignación: Posibilitan que un sujeto que fue asignado a un grupo en un determinado paso del proceso sea reasignado a otro grupo en un paso posterior, si ello perfecciona el criterio de selección. El proceso termina cuando no quedan sujetos cuya reasignación permita la optimización del resultado que se quiere conseguir. Dentro de estos métodos se tienen:
 - o El método K-Medidas
 - o El Quick-Clúster análisis
 - o El método de Forgy
 - o El método de las nubes dinámicas
- Métodos de búsqueda de la densidad: Dentro de estos métodos están los que brindan un acercamiento tipológico y una aproximación de mayor probabilidad.

En el primer tipo, los grupos son formados buscando zonas donde hay mayor concentración de sujetos. Entre ellos tenemos:

- o El análisis modal de Wishart

- El método Taxamp
- El método de Fortin

En el segundo tipo se inicia a partir de la premisa de que las variables se guían por una ley de probabilidad en el que los parámetros varían de un grupo a otro. El objetivo es encontrar los individuos que pertenecen a la misma distribución. Entre los métodos de este tipo el método de las combinaciones de Wolf es la que destaca.

- Métodos directos: Indaga la clasificación de manera simultánea a individuos y variables. En este método destaca:
 - Clusterización usando cuadros (Bloc Clustering)
- Métodos de reducción de dimensiones: Averigua factores en el espacio de los individuos, cada factor pertenece a un grupo. Como Análisis Factorial tipo Q es también conocido.

Pedret et al. (2003) indica que para la agrupación de extensos conjuntos de datos como pueden ser clasificaciones de estudiantes, consumidores, compradores, etc.; es recomendable aplicar estos métodos, de acuerdo a sus propios comportamientos y peculiaridades, posturas, etc.

Pero cuando la idea de la cantidad de grupos que puede estar dentro los datos no es obtenida aún en la fase de exploración, es necesario el uso de un método no jerárquico para que la clasificación que se debe obtener final pueda ser obtenida y se confirme la pertenencia de los grupos que fueron encontrados; la cantidad de grupos también se puede obtener por medio de un método jerárquico.

Según Gallardo (s.f.) menciona algunos cuidados que se debe tomar en cuenta sobre los métodos clúster:

- a) Algunos de los métodos de análisis clúster son procedimientos que, en la mayoría de los casos, no están soportados por un sustento doctrinal estadístico teórico. En otras palabras, la mayor parte de los métodos son hallazgos.
- b) Los diferentes procedimientos de clúster pueden crear distintas soluciones sobre el mismo conjunto de datos. Una razón para ello radica en el hecho de que los métodos clúster fueron desarrollados a partir de diferentes disciplinas que han

dado principio a diferentes reglas de formación de grupos. De esta manera lógica, la existencia de estas técnicas es necesaria porque es necesario conocer qué método es capaz de producir los grupos más homogéneos de manera natural en los datos [31].

En la presente tesis, se usará el método K-medidas o K-means por medio de un algoritmo en el programa Python en Google Colab que se mostrará más adelante.

1.6.2.2 Método de Elbow

El método utiliza los valores de la inercia obtenidos tras aplicar el K-medidas a diferente número de clústeres (desde 1 a N clusters), siendo la inercia la suma de las distancias al cuadrado de cada objeto Cluster a su centroide.

Una vez obtenidos los valores de la inercia tras aplicar el K-medidas de 1 a N clústeres, representamos en una gráfica lineal la inercia respecto del número de clústeres. En la Figura 1.9 se puede apreciar cómo es la gráfica que se aplica por medio del método de Elbow [32].

En la Figura 1.9 se puede apreciar un ejemplo de cómo es gráfica obtenida por el método de Elbow, pero halla por medio del programa R.

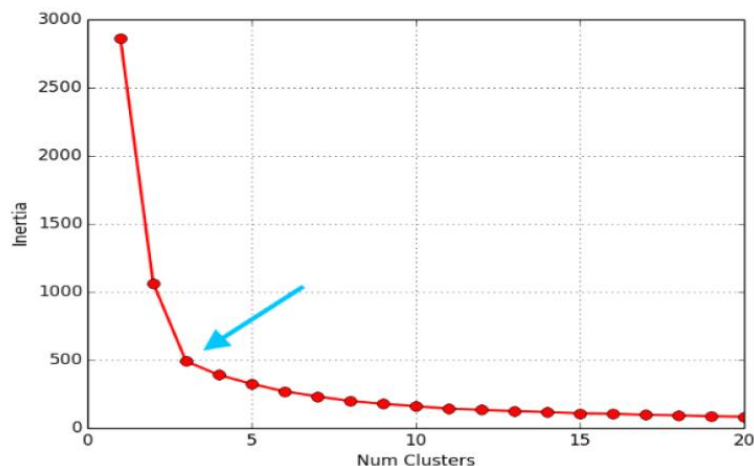


Figura 1.9 Ejemplo de gráfica obtenido por el método de Elbow por medio del programa R [32]

1.6.2.3 Las 4 regresiones

En la presente tesis después de la clusterización se hará un análisis de 4 regresiones que permitirá realizar las predicciones de las cuales se dará su definición.

- a) Regresión Logística: Sirve solamente para la clasificación, pero no para la regresión. El algoritmo se implementa de manera muy simple, pero su convergencia es demasiado compleja si los datos que van a ser tratados no son separables linealmente. Este clasificador es lineal binario, sin embargo, es la técnica OvR (One versus Rest) la que usa.

Estadísticamente, es un modelo de regresión donde la variable dependiente es categórica, aquella que puede tomar valores fijos o una cantidad de valores posibles. Estas variables dependientes son aquellas clases de objetivos en el que se quiere realizar la predicción mientras que las variables independientes son las de diferentes cualidades que constituyen el conjunto de datos y que va a ser usado en el modelo que se va a mostrar en la metodología.

- b) Regresión K-NN: Llamado el algoritmo K-Nearest Neighbors, este algoritmo es uno de los de mayor popularidad para el reconocimiento de patrones y correlaciones. Un ejemplo para esta regresión es el modelo financiero para prever cómo estarán los mercados para la detección de spam, o también para el área de la medicina con la finalidad de saber si un paciente que ha ingresado por un infarto al corazón es probable que sufra otro a su historial médico.

La implementación de este algoritmo es muy sencilla de implementar para clasificación. Elegir este algoritmo puede ser una buena idea, si no tenemos la una buena primera idea si no se cuenta con el conocimiento de cómo están distribuidos los datos que se quieren clasificar. No asume cómo están distribuidos los datos, algo que en el mundo real así ocurre, ya que la mayoría de los datos no obedecen a una distribución que está de manera teórica que, pero en el caso de la regresión logística, estas si obedecen esa distribución. En otras palabras, escoger este algoritmo puede ser una excelente elección siempre y cuando no se conozca la manera en la que están distribuidos que se van a clasificar.

Este método es considerado vago en la manera que apenas usa los datos de prueba para la construcción de un modelo de manera general y usarlo para la realización de la clasificación, pues, en la fase de testeo, son bien usados los datos a usar. La razón es porque la clasificación, en cada uno de los datos que se dispone en el conjunto usado para el entrenamiento, está basada en la averiguación de la similitud en las características de cada dato usado, y que, al tener que usar el conjunto en su totalidad para comparar uno por uno, se requiere adicionalmente una memoria de almacenamiento inmejorable.

Otra característica de este método es que no tiene parámetros, en otras palabras, el método no hace el análisis de manera hipotética de cómo los datos están distribuidos, supone que los mismos datos con los que se trabaja son el mejor modelo de datos.

Para cada nuevo objeto observado, la clasificación está basada en la asignación a la clase de mayor votación de sus k vecinos cercanos. K es el valor que determina la cantidad de vecinos de mayor cercanía que tengan las características de la misma similitud con la finalidad de que la comparación se realice y la elección que se lleve a cabo.

- c) Regresión SVM: SVM significa Support Vector Machine. Este algoritmo puede ser considerado como una ampliación de Perceptrón. De mayor importancia es saber de qué forma este algoritmo obtiene el hiperplano más excelente para una clasificación binaria, un hiperplano que se encarga de la separación de aquellos ejemplos que pertenecen a una u otra clase.
- d) Regresión Decision Tree: El uso de Arboles de decisiones pueden ser aplicados en distintas áreas, áreas como la procedencia y clasificación de texto; en el área de la medicina se puede usar para la detección de enfermedades como el cáncer, problemas de salud en el corazón, enfermedades neurodegenerativas o ver si es necesario la hospitalización de un paciente que fue diagnosticado con dengue, en el área de la economía puede ser aplicado para ver cómo se comporta la bolsa de valores por medio de los mercados de valores.

Este algoritmo tiene reglas de clasificación que son fáciles de comprender para las personas lo que lo más efectivo y lo que más destaca del algoritmo. De la estructura de un árbol es la idea de donde viene el nombre este algoritmo, ya que la estructura del árbol se compone una raíz y unos nodos; en los nodos se puede apreciar el despliegue hacia debajo de 2 o más ramas de izquierda a derecha.

Un árbol de decisión es realizado por medio del agrupamiento de datos que se basa en los valores de los atributos de los datos recibidos. De acuerdo al atributo que es el más distinguido de los demás se realiza un división y de manera recursiva se aplica este proceso es aplicado hasta que todos los datos de un subconjunto que esté siendo tratado pertenezca a la misma clase [33].

De cada una de las 4 regresiones se obtendrá una matriz de confusión.

1.6.2.4 Matriz de confusión

Esta matriz es una herramienta que ayuda a saber cuál es el rendimiento del algoritmo. Después de que el entrenamiento y la validación de los datos que se van a analizar para que sean testeados, la matriz mostrará el siguiente testeo que se puede ver en la Figura 1.10 que son el verdadero positivo, el verdadero negativo, falso positivo y falso negativo.

	Positivo	Negativo
Positivo	verdadero positivo	falso positivo
Negativo	falso negativo	verdadero negativo

Figura 1. 10 Matriz de confusión [34]

- a) Verdaderos positivos: Son todas las observaciones que fueron clasificadas como pertenecientes a una clase y acertó.
- b) Verdaderos negativos: Son los valores que no pertenecían a los verdaderos positivos y que acertaron.
- c) Falsos positivos: Son las observaciones no pertenecientes a una clasificación, pero que se consideraron pertenecientes a ellas, se consideran positivos, pero no lo son.

- d) Falsos negativos: Son esclarecimientos pertenecientes a una clasificación, pero no son considerados como tales, por lo que se les considera como negativos.

Después de obtener la matriz de confusión se puede calcular el ratio de precisión y el ratio de error.

- a) Exactitud: Este porcentaje sirve para obtener el porcentaje de clasificaciones que ha sido clasificado correctamente. Cuanto más cercano a 1 sea la exactitud, mejor será.

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN}$$

Ecuación 1.1 Ecuación para la exactitud [34]

$$\text{Exactitud promedio} = \frac{1}{c} \sum_{i=1}^c \frac{VP_i + VN_i}{VP_i + VN_i + FP_i + FN_i}$$

Ecuación 1.2 Ecuación para calcular la exactitud promedio [34]

C es el número de clases cuando se realiza una clasificación multiclase.

El ratio de error se obtiene cuando el 100% de probabilidad, en otras palabras 1, se resta con el resultado del ratio de la exactitud.

- b) Precisión: La precisión mide la habilidad de un clasificador de no etiquetar como positivo una observación que se debe considerar como negativa. Mejor es la precisión si este es mayor [34].

$$\text{Sensibilidad} = \frac{VP}{VP + FN}$$

Ecuación 1.3 Ecuación para la sensibilidad [34]

$$\text{Sensibilidad promedio} = \frac{1}{c} \sum_{i=1}^c \frac{VP_i}{VP_i + FN_i}$$

Ecuación 1.4 Ecuación para la sensibilidad promedio [34]

- c) Precisión: La precisión mide la habilidad de un clasificador de no etiquetar como positivo una observación que se debe considerar como negativa. Mejor es la precisión si este es mayor [34].

$$\text{Precisión} = \frac{VP}{VP + FP}$$

Ecuación 1.5 Ecuación para la precisión [34]

$$\text{Precisión promedio} = \frac{1}{c} \sum_{i=1}^c \frac{VP_i}{VP_i + FP_i}$$

Ecuación 1.6 Ecuación para la precisión promedio [34]

1.7 Teorema de Manning

Para analizar el problema, debemos tomar en cuenta el teorema de Manning, ya que el agua circula por la acción de la gravedad y sin ninguna presión, ya que para este teorema tiene que ver con la sección del tubo y el agua solo abarca una sección del colector. Ahora se procederá a dar a conocer las fórmulas del teorema de Manning con el fin de entender cómo funciona la circulación del agua en los colectores.

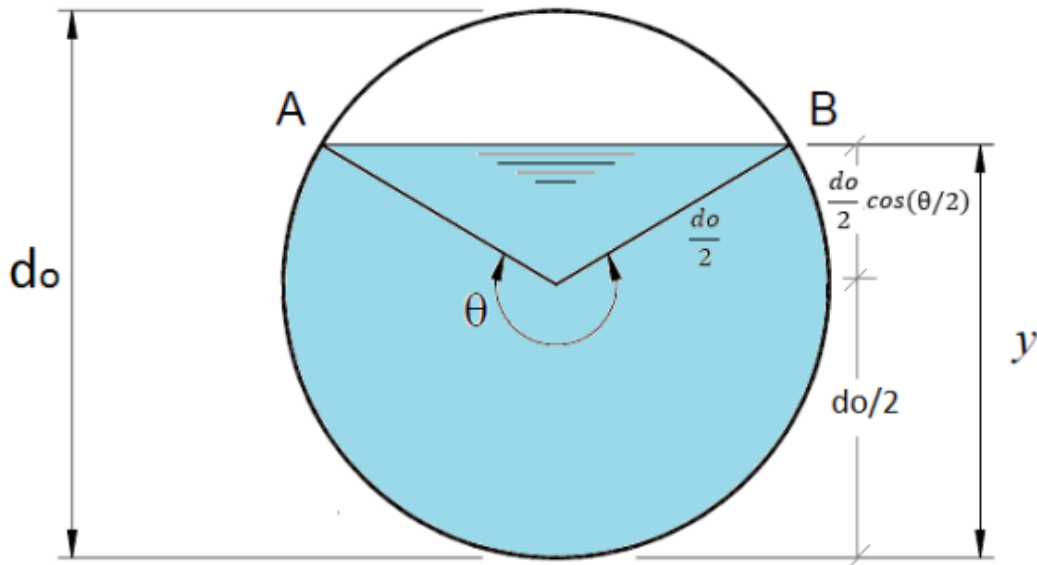


Figura 1. 11 Sección de tubo por el que pasa el flujo del agua [33]

Tirante (y):

$$y = \frac{d_0}{2} * (1 - \cos(\frac{\theta}{2}))$$

Ecuación 1.7 Ecuación para el tirante [35]

Área (A):

$$A = \frac{d_0^2}{8} * (\theta - \sin(\theta))$$

Ecuación 1.8 Área mojada del ducto [33]

Radio hidráulico (R_h):

$$R = \frac{d_0}{4} * (1 - \frac{\sin(\theta)}{\theta})^{\frac{2}{3}}$$

Ecuación 1.9 Radio hidráulico del área mojada [35]

Perímetro mojado (P):

$$P = \frac{d_0}{2} * \theta$$

Ecuación 1.10 Perímetro que ocupa el área mojada [33]

Caudal (Q):

$$Q = \frac{1}{2^{\frac{13}{3}}} * \frac{S^{\frac{1}{2}}}{n} * \frac{(\theta - \sin(\theta))^{\frac{5}{3}}}{\theta^{\frac{2}{3}}} * \theta^{\frac{8}{3}}$$

Ecuación 1.11 Caudal del área mojada [35]

Velocidad crítica (V_c):

$$V_c = 6 * \sqrt{g * R_h}$$

Ecuación 1.12 Velocidad crítica [35]

Donde:

n es rugosidad.

S es pendiente.

g es gravedad [35].

CAPÍTULO II

METODOLOGÍA

En el presente capítulo se detalla el procedimiento seguido para la predicción de fallas en el sistema de colectores de aguas servidas del distrito de San Juan de Lurigancho. Como se mencionó en el marco teórico, el método utilizado fue el de la minería de datos (Data Mining) por medio de la herramienta de Machine Learning, ya que la cantidad de datos que se cuenta con una gran cantidad de datos. A continuación, se presenta el esquema general del trabajo, cuyos pasos se detallan en las siguientes secciones.

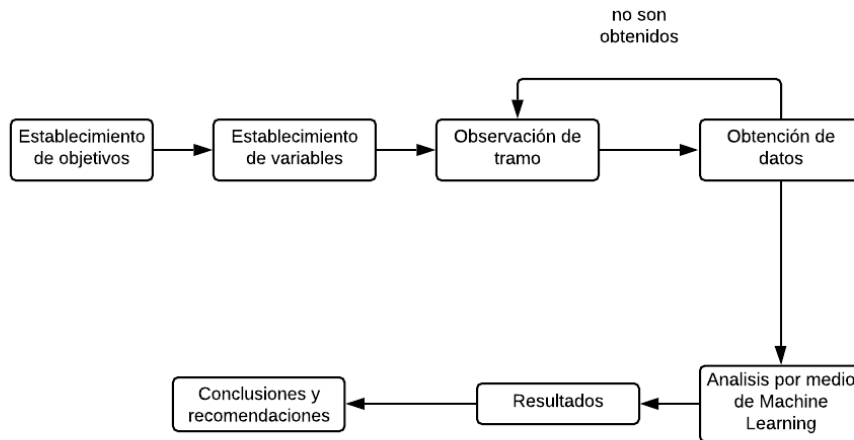


Figura 2. 1 Diagrama de metodología

2.1 Establecimiento de variables

Para el desarrollo de la metodología es necesario tener en cuenta las variables que se van a utilizar, las cuales son longitud, edad, diámetro, pendiente y el material del colector, así como profundidad de aguas arriba, profundidad de aguas abajo, cuenca y la zona del

sistema de aguas servidas. A continuación, se muestra una tabla resumen con la definición de cada variable, los valores encontrados y su fuente de información.

Variable	Definición	Valores	Fuente y forma de recolección de información
Longitud del colector	Es el tramo que se debe tomar en cuenta para el análisis.	m	[37]
Edad del Colector	Es la edad que tiene el colector después de su colocación hasta la actualidad. Esta variable se mide en años.	Meses	[37]
Diámetro del colector	Es la distancia circunferencial del colector. Esta es una variable de bajo alcance. Las tuberías que son de un diámetro más pequeño son las que tienen mayor tendencia a obstruirse, mientras que las tuberías más grandes tienen mayor tendencia a sufrir un colapso.	mm para el caso 1 y m para el caso 2	[37]
Pendiente del colector	La pendiente es la inclinación del colector, en este caso lineal. Los colectores que tienen mayor tendencia a la obstrucción son los colectores de pendiente baja y los que poseen una pendiente baja son las que presentan una mayor velocidad y una mayor tendencia al desgaste. Esta se mide como un porcentaje en relación con el descenso de la tubería a medida que avanza.	Porcentaje	[37]
Material del colector	En el capítulo 1 se mencionó cada uno de los materiales de las que son hechas las tuberías para el servicio de alcantarillado, pero el tramo a analizar va a ser una tubería de concreto. En este caso, algunos autores, como Chughtai y Zayed, han usado valores numéricos en forma	Concreto	[37,40]

	<p>del coeficiente de rugosidad de la dotación, pero no brinda información sobre la tendencia de una tubería a deteriorarse o a colapsarse, y por lo tanto se debe tener cuidado a la hora de usar esa variable. (definición)</p> <p>El coeficiente de rugosidad para el concreto liso es de 0.01; para el concreto norma, 0.013; para el PVC, 0.01; para el fierro fundido, 0.01, y para la arcilla vitrificada, 0.013.</p>		
Profundidad	<p>Altura a la que un colector se encuentra. Una tubería que se encuentra a menor profundidad está más sujeta las presiones que vienen de la superficie o al daño que causa las raíces cuando penetran en la tubería. Una tubería que se encuentra a mayor profundidad está más sujeta a la presión externa provista por el suelo circundante y la filtración del agua subterránea.</p>	m	[37]
Caudal	Es el flujo por la que pasa el agua	m ³ /s	[35]
Tirante	<p>Es la altura que ocupa el agua en ducto, por diseño, el tirante debe ser menor o igual del 75% del diámetro, pero si la velocidad crítica es menor que la velocidad encontrada, el tirante debe ser menor o igual al 50% del diámetro.</p> <p>Con esta variable se puede calcular el área mojada, el perímetro mojado el Caudal, la velocidad crítica y el ángulo para que genera el tirante como se puede ver en la Figura 1.11 del capítulo 1.</p> <p>La empresa SEDAPAL informó que la rugosidad de Manning que utilizan es de 0.009</p>	m	[35,36]
Velocidad	La velocidad, según la NORMA TÉCNICA OS-010, si es conducción por efecto de	m/s	[36]

	gravedad, la velocidad mínima no debe producir depósitos y erosiones, por lo que no debe ser menor de 0.6 m/s. La velocidad máxima (crítica) admisible para tubos de concreto es de 3 m/s; en los tubos de asbesto-cemento y PVC acero, 5 m/s.		
Número de actividades	Se debe tomar en cuenta las actividades que se realizan en un determinado lugar, ya que estas pueden ser las causas de múltiples atoros	veces	Observación
Número de fallas	En este caso, se debe tomar en cuenta que el número de fallas que sufre un colector puede ocasionar que la condición del colector vaya perdiéndose	veces	Observación

Tabla 2. 1 Tabla de definición de las variables junto con sus unidades y fuentes

2.2 Selección de tramos

Después de tener establecidas las variables, Se tomó la decisión de analizar los tramos secundarios.

2.3 Obtención de datos

Para la obtención de datos del tramo seleccionado, se envió una carta a la empresa SEDAPAL solicitando los siguientes datos:

- Tirante en el colector.
- Diámetro del colector.
- Caudal en el colector.
- Material del Colector.

2.4 Análisis por medio de Machine Learning

En la presente tesis se usó el método de Machine Learning por medio del método de clusterización, se realizó por medio del siguiente algoritmo en Python aprendido en el Curso de Especialización de Ciencia de Datos orientado al Mantenimiento Predictivo de la Escuela de Formación Profesional Ingenium en las clases 7, 9 y 10 [38] [39]:

```
## importamos librerías
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

##importamos la data
dataset=pd.read_excel('colectores medidas2.xlsx')
dataset

## X son las columnas predictoras para clusterizar
X = dataset.iloc[:, [0,6]].values

## Usamos el método de Elbow para encontrar el número óptimo de clústeres.
# sugerencia de cantidad de cluster
from sklearn.cluster import KMeans
wcss = []

for i in range(1, 9):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 9), wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')

## Entrenamiento por medio del modelo K-Means (K-medidas) en los datos importados
kmeans = KMeans(n_clusters = 3, init = 'k-means++', random_state = 42)
y_kmeans = kmeans.fit_predict(X)

## Se visualizan los clústeres
plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 100, c = 'red', label = 'Normal')
plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 100, c = 'green', label = 'Posible atoro')
plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 100, c = 'blue', label = 'Atoro')
# plt.scatter(X[y_kmeans == 3, 0], X[y_kmeans == 3, 1], s = 100, c = 'cyan', label = 'Cluster 3')
# plt.scatter(X[y_kmeans == 4, 0], X[y_kmeans == 4, 1], s = 100, c = 'magenta', label = 'Cluster 5')

plt.title('Clusters de colectores')
plt.xlabel('Tiempo de duración del colector en mese')
plt.ylabel('duración del colector')
plt.legend(loc=4)
```



```

plt.grid()

## juntando df
y_kmeans = y_kmeans.reshape(len(y_kmeans),1)
ydf = pd.DataFrame(y_kmeans)
ydf.columns=['k-means']

total=pd.concat([dataset,ydf],axis=1)
total

total.dtypes

total.head()

```

Luego de hacer la clusterización se utilizó los algoritmos relacionados a las 4 regresiones antes mencionadas en el capítulo 1, los cuales son:

a) Regresión logística:

```

total

total.iloc[:, :-1].corr()

import seaborn as sns

corr = total.iloc[:, :-1].corr()
ax = sns.heatmap(
    corr,
    vmin=-1, vmax=1, center=0,
    cmap=sns.diverging_palette(20, 220, n=200),
    square=True
)
ax.set_xticklabels(
    ax.get_xticklabels(),
    rotation=45,
    horizontalalignment='right'
);

## Escalado de variables
from sklearn.preprocessing import StandardScaler
sc_X = StandardScaler()
total.iloc[:, :-1 ] = sc_X.fit_transform(total.iloc[:, :-1 ])

total.head()

## matriz de características
X2 = total.iloc[:, :-1 ].values
y = total.iloc[:, -1].values

## Reducir la dimensión del dataset con ACP a 10 componentes
from sklearn.decomposition import PCA
pca = PCA(n_components = 2)
X2 = pca.fit_transform(X2)
explained_variance = pca.explained_variance_ratio_

```

```

## Splitting the dataset into the Training set and Test set
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X2, y, test_size = 0.20, random_state = 0)

## Feature Scaling
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

## Algoritmo Logistic regression
# Training the Logistic Regression model on the Training set
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 0)
classifier.fit(X_train, y_train)

# Predicting the Test set results
# y_pred son los valores teóricos

y_pred = classifier.predict(X_test)

y_pred

X_test

## Haciendo la matriz de confusión
# y_test son valores reales

from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)

# Visualising the Training set results
from matplotlib.colors import ListedColormap
X_set, y_set = X_train, y_train
X1, X3 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                    np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X3, classifier.predict(np.array([X1.ravel(), X3.ravel()]).T).reshape(X1.shape),
            alpha = 0.75, cmap = ListedColormap(('red', 'green', 'blue')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X3.min(), X3.max())
for i, j in enumerate(np.unique(y_set)):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
                c = ListedColormap(('red', 'green', 'blue'))(i), label = j)
plt.title('LR (Training set)')
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.legend()

# Visualising the Test set results
from matplotlib.colors import ListedColormap
label=['Seguro', 'Intermedio', 'Crítico']
X_set, y_set = X_test, y_test
X1, X3 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                    np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X3, classifier.predict(np.array([X1.ravel(), X3.ravel()]).T).reshape(X1.shape),
            alpha = 0.75, cmap = ListedColormap(('red', 'green', 'blue')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X3.min(), X3.max())
for i, j in zip(np.unique(y_set), np.unique(y_set), label):

```

```

plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
            c = ListedColormap(('red', 'green', 'blue'))(i), label = (str(j)+ ': '+lab))
plt.title('LR (Test set)')
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.legend()

```

b) Regresión K-NN:

```

total

total.iloc[:, :-1].corr()

import seaborn as sns

corr = total.iloc[:, :-1].corr()
ax = sns.heatmap(
    corr,
    vmin=-1, vmax=1, center=0,
    cmap=sns.diverging_palette(20, 220, n=200),
    square=True
)
ax.set_xticklabels(
    ax.get_xticklabels(),
    rotation=45,
    horizontalalignment='right'
);

from sklearn.preprocessing import StandardScaler
sc_X = StandardScaler()
total.iloc[:, :-1 ] = sc_X.fit_transform(total.iloc[:, :-1 ])

total.head()

# matriz de características

X2 = total.iloc[:, :-1 ].values
y = total.iloc[:, -1].values

# Reducir la dimensión del dataset con ACP a 10 componentes
from sklearn.decomposition import PCA
pca = PCA(n_components = 2)
X2 = pca.fit_transform(X2)
explained_variance = pca.explained_variance_ratio_

# Splitting the dataset into the Training set and Test set
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X2, y, test_size = 0.20, random_state = 0)

# Feature Scaling
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Training the K-NN model on the Training set
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)

```

```

classifier.fit(X_train, y_train)

# Predicting the Test set results
y_pred = classifier.predict(X_test)

# Making the Confusion Matrix
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)

# Visualising the Training set results
from matplotlib.colors import ListedColormap
label=['Seguro', 'Intermedio', 'Crítico']
X_set, y_set = X_train, y_train
X1, X3 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                    np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X3, classifier.predict(np.array([X1.ravel(), X3.ravel()]).T).reshape(X1.shape),
             alpha = 0.75, cmap = ListedColormap(('red', 'green','blue')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X3.min(), X3.max())
for i, j in enumerate(np.unique(y_set)):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
               c = ListedColormap(('red', 'green','blue'))(i), label = (str(j)+ ': '+label))
plt.title('K-NN (Training set)')
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.legend()

# Visualising the Test set results
from matplotlib.colors import ListedColormap
X_set, y_set = X_test, y_test
X1, X3 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                    np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X3, classifier.predict(np.array([X1.ravel(), X3.ravel()]).T).reshape(X1.shape),
             alpha = 0.75, cmap = ListedColormap(('red', 'green','blue')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X3.min(), X3.max())
for i, j in zip(np.unique(y_set), np.unique(y_set), label):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
               c = ListedColormap(('red', 'green','blue'))(i), label = j)
plt.title('K-NN (Test set)')
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.legend()

```

c) Regresión SVM:

```

total

total.iloc[:, :-1].corr()

import seaborn as sns

corr = total.iloc[:, :-1].corr()
ax = sns.heatmap(
    corr,
    vmin=-1, vmax=1, center=0,
    cmap=sns.diverging_palette(20, 220, n=200),

```

```

square=True
)
ax.set_xticklabels(
    ax.get_xticklabels(),
    rotation=45,
    horizontalalignment='right'
);

from sklearn.preprocessing import StandardScaler
sc_X = StandardScaler()
total.iloc[:, :-1] = sc_X.fit_transform(total.iloc[:, :-1])

total

# separando entre variables predictoras y para predecir

X2 = total.iloc[:, :-1].values
y = total.iloc[:, -1].values

# Reducir la dimensión del dataset con ACP a 10 componentes
from sklearn.decomposition import PCA
pca = PCA(n_components = 2)
X2 = pca.fit_transform(X2)
explained_variance = pca.explained_variance_ratio_

# Splitting the dataset into the Training set and Test set
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X2, y, test_size = 0.20, random_state = 0)

# Feature Scaling
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Training the SVM model on the Training set
from sklearn.svm import SVC
classifier = SVC(kernel = 'linear', random_state = 0)
classifier.fit(X_train, y_train)

# Predicting the Test set results
y_pred = classifier.predict(X_test)

y_pred

# Making the Confusion Matrix
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)

# Visualising the Training set results
from matplotlib.colors import ListedColormap
label=['normal','alerta','emergencia']
X_set, y_set = X_train, y_train
X1, X3 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                    np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X3, classifier.predict(np.array([X1.ravel(), X3.ravel()]).T).reshape(X1.shape),
            alpha = 0.75, cmap = ListedColormap(('red', 'green', 'blue')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X2.min(), X2.max())

```

```

for i, j, lab in zip(np.unique(y_set).np.unique(y_set).label): #se cambia el enumerate porque se está
utilizando un label como texto
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
                c = ListedColormap(('red', 'green', 'blue'))(i), label = (str(j)+': '+lab))
plt.title('SVM (Training set)')
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.legend()

# Visualising the Test set results
from matplotlib.colors import ListedColormap
label=['Seguro', 'Intermedio', 'Crítico']
X_set, y_set = X_test, y_test
X1, X3 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                    np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X3, classifier.predict(np.array([X1.ravel(), X3.ravel()]).T).reshape(X1.shape),
             alpha = 0.75, cmap = ListedColormap(('red', 'green', 'blue')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X2.min(), X2.max())
for i, j, lab in zip(np.unique(y_set).np.unique(y_set).label):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
                c = ListedColormap(('red', 'green', 'blue'))(i), label = (str(j)+': '+lab))
plt.title('SVM (Test set)')
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.legend()

```

d) Regresión Decision Tree:

```

total

total.iloc[:, :-1].corr()

import seaborn as sns

corr = total.iloc[:, :-1].corr()
ax = sns.heatmap(
    corr,
    vmin=-1, vmax=1, center=0,
    cmap=sns.diverging_palette(20, 220, n=200),
    square=True
)
ax.set_xticklabels(
    ax.get_xticklabels(),
    rotation=45,
    horizontalalignment='right'
);

# Escalado de variables

from sklearn.preprocessing import StandardScaler
sc_X = StandardScaler()
total.iloc[:, :-1 ] = sc_X.fit_transform(total.iloc[:, :-1 ])

# la matriz de características

X2 = total.iloc[:, :-1 ].values
y = total.iloc[:, -1 ].values

```

```

# Reducir la dimensión del dataset con ACP a 10 componentes
from sklearn.decomposition import PCA
pca = PCA(n_components = 2)
X2 = pca.fit_transform(X2)
explained_variance = pca.explained_variance_ratio_

# Splitting the dataset into the Training set and Test set
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X2, y, test_size = 0.20, random_state = 0)

# Feature Scaling
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

# Training the Decision Tree Classification model on the Training set
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier(criterion = 'gini', random_state = 0, max_depth=2)
classifier.fit(X_train, y_train)

# Predicting the Test set results
y_pred = classifier.predict(X_test)

# Making the Confusion Matrix
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)

# Visualising the Training set results
from matplotlib.colors import ListedColormap
X_set, y_set = X_train, y_train
X1, X3 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                    np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X3, classifier.predict(np.array([X1.ravel(), X3.ravel()]).T).reshape(X1.shape),
            alpha = 0.75, cmap = ListedColormap(('red', 'green', 'blue')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X3.min(), X3.max())
for i, j in enumerate(np.unique(y_set)):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
                c = ListedColormap(('red', 'green', 'blue'))(i), label = j)
plt.title('Decision Tree Classification (Training set)')
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.legend()

# Visualising the Test set results
from matplotlib.colors import ListedColormap
label=['Seguro', 'Intermedio', 'Crítico']
X_set, y_set = X_test, y_test
X1, X3 = np.meshgrid(np.arange(start = X_set[:, 0].min() - 1, stop = X_set[:, 0].max() + 1, step = 0.01),
                    np.arange(start = X_set[:, 1].min() - 1, stop = X_set[:, 1].max() + 1, step = 0.01))
plt.contourf(X1, X3, classifier.predict(np.array([X1.ravel(), X3.ravel()]).T).reshape(X1.shape),
            alpha = 0.75, cmap = ListedColormap(('red', 'green', 'blue')))
plt.xlim(X1.min(), X1.max())
plt.ylim(X3.min(), X3.max())
for i, j in zip(np.unique(y_set), np.unique(y_set), label):
    plt.scatter(X_set[y_set == j, 0], X_set[y_set == j, 1],
                c = ListedColormap(('red', 'green', 'blue'))(i), label = (str(j)+' '+label))

```

```
plt.title('Decision Tree Classification (Test set)')
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.legend()
```

Primero se debe tomar en realizar un Excel con todos los datos obtenidos para importarlos al código Python en Google Colab y hacer la clusterización, segundo se pone en marcha los códigos mencionados en este capítulo.

De los 4 regresiones se obtendrá una matriz de confusión.

2.5 Resultados

Después de obtenidos los resultados del Machine Learning aplicando la clusterización y las 4 regresiones, se realizó un análisis de acuerdo a esos resultados en función de los siguiente:

- Edad
- Profundidad
- Diámetro
- Tirante
- Profundidad
- Material (Rugosidad de Manning)
- Números de fallas
- Número de actividades

CAPÍTULO III

RESULTADOS

En el presente capítulo se presentan los resultados del estudio de acuerdo a la metodología propuesta.

3.1 Variables establecidas

Las variables que se llegaron a establecer estarán separadas en 2 escenarios que son escenario 1 y escenario 2:

Para el escenario 1 las variables son:

- Longitud del colector
- Edad del colector
- Diámetro del colector
- Pendiente del colector
- Profundidad
- Número de fallas
- Número de fallas acumulativo

Para el escenario 2 las variables son:

- Diámetro
- Material del colector (Rugosidad de Maning)
- Ángulo
- Área mojada (m²)
- Perímetro mojado (m)
- Radio hidráulico (m)
- Pendiente

- Caudal (L/s)
- Velocidad (m/s)
- Velocidad crítica (m/s)
- Número de actividades
- Número de actividades acumulativo

Como la cantidad de datos es muy grande, para ver los valores de las variables mencionadas, se pueden ver en el archivo modelo de prueba (1).

3.1.1 Selección de tramo

Se hizo una observación en el distrito de SJL por medio de Google Earth y una visita por medio de recorrido en auto y se decidió hacer el análisis de los tramos secundarios, ya que en esos tramos es donde más ocurre los incidentes.

3.2 Machine Learning, Clusterización y Los 4 Algoritmos para el Análisis

Para este caso se tuvo que analizar 2 escenarios, en ellos se puede las matrices de confusión, el valor de los ratios mencionados de cada una de las regresiones mencionadas en el capítulo 1, la única diferencia es que, en lugar de dos clasificaciones, como el positivo y el negativo, se usarán 3 para clasificar las condiciones en los 2 escenarios que estarán clasificados por números:

- Para el escenario 1: 0 es Seguro, 1 es Intermedio y 2 es Crítico.
- Para el escenario 2: 0 es Normal, 1 es Probabilidad de atoro y 2 es Atoro.

Por lo tanto, el cálculo de la sensibilidad no será necesaria, ya que ese ratio es para el caso de que la clasificación sea binaria como se muestra en el capítulo 1.

3.2.1 Condición del colector según el diámetro, edad, profundidad y pendiente

Como modelo de prueba se estableció un cuadro de prueba usando algunos de los valores faltantes con tal de desarrollar un avance que se puede apreciar en la tabla 1 del Anexo 3.

Primero se debe aplicar el método de Elbow para determinar la cantidad de clústeres antes de hacer la clusterización, en la Tabla 1 del Anexo 3.

En este escenario se realizará la clusterización por medio de tres de los valores: el tiempo, el número de fallas acumuladas del colector y el número de fallas, ya que la cantidad de fallas por las que pasa los colectores determina la condición de estos mismos. Para las predicciones se tomará todo.

En la Tabla 1 del Anexo 3 se puede apreciar la tabla con los valores a usar.

En la Figura 3.1 se puede apreciar la gráfica del método de Elbow para determinar el número de clústeres.

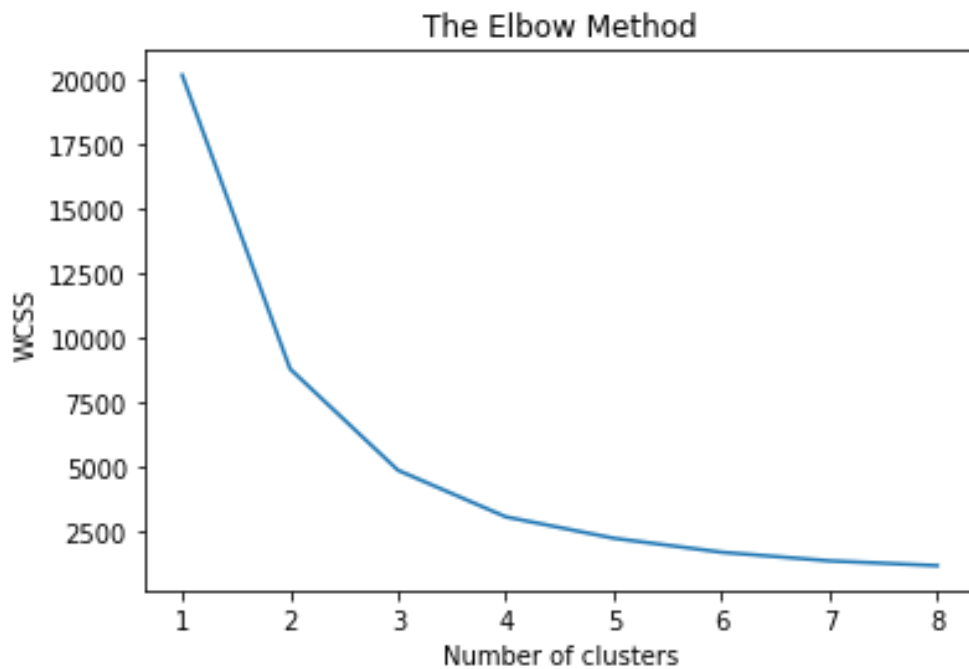


Figura 3. 1 Gráfica de método de Elbow del Escenario 1.

Se analizó la figura 3.1, es recomendable hacer de entre 2 a 3 clusters, pero en este caso se optó por hacer 3 clústeres.

Viendo la Figura 3.2, se tomó la decisión de hacer 3 clústeres. En la siguiente gráfica se puede apreciar la clusterización.

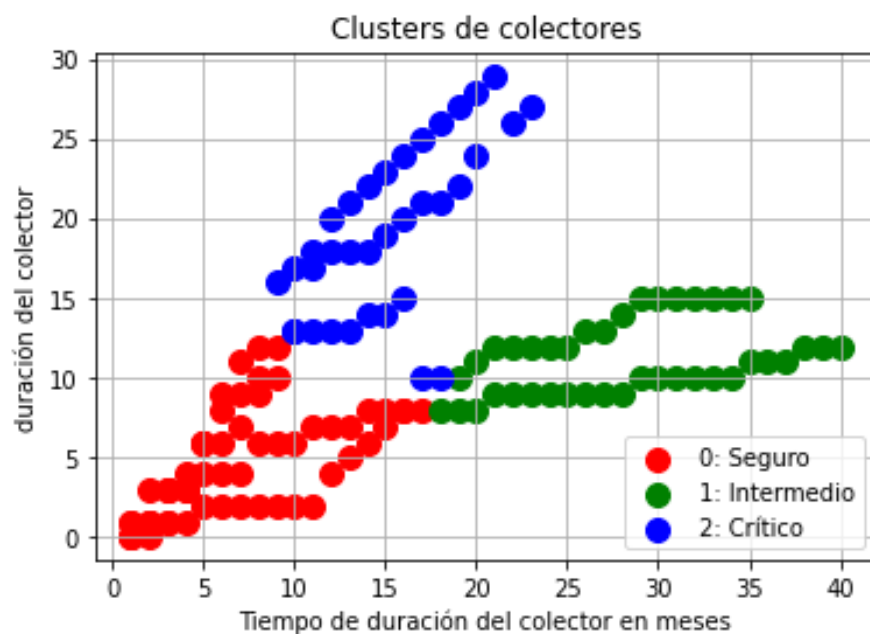


Figura 3. 2 Gráfica de Clusterización del Escenario 1.

En la Tabla 1 del Anexo 2 se puede apreciar la clusterización de los datos.

En la Figura 3.3 se puede apreciar una tabla abreviada del resultado de la clusterización del escenario 1.

	Diametro (mm)	longitud (m)	Pendiente	Profundidad (m)	Tiempo (años)	Número de fallas (acumulativo)	Número de fallas	k-means
0	762	1000.0	0.01172	9	1	1	1	0
1	762	1000.0	0.01172	9	2	3	2	0
2	762	1000.0	0.01172	9	3	3	0	0
3	762	1000.0	0.01172	9	4	3	0	0
4	762	1000.0	0.01172	9	5	4	1	0
...
129	1220	1800.0	0.00111	14	12	13	0	2
130	1220	1800.0	0.00111	14	13	13	0	2
131	1220	1800.0	0.00111	14	14	14	1	2
132	1220	1800.0	0.00111	14	15	14	0	2
133	1220	1800.0	0.00111	14	16	15	1	2

Figura 3. 3 Resultados de la clusterización de k-mean del Escenario 1.

En la Figura 3.4 se mostrará la gráfica de las correlaciones.

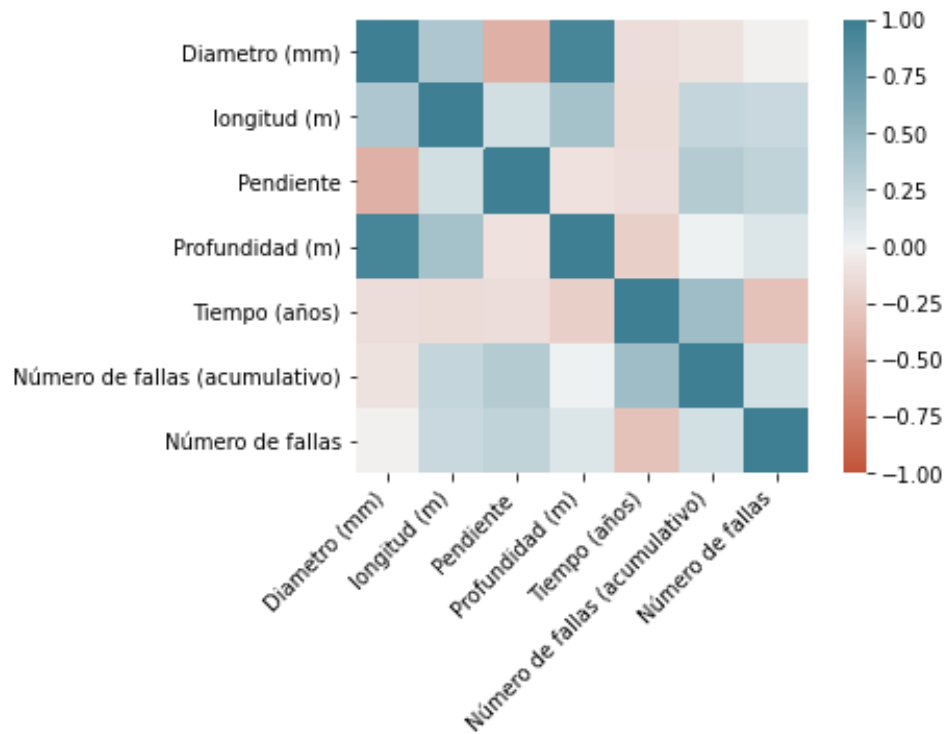


Figura 3. 4 Gráfica de la correlación de datos del Escenario 1.

Ahora se mostrará los resultados de cada una de las regresiones.

a) Regresión logística: En las Figuras 1a y 1b del Anexo 4 se puede apreciar las gráficas de la Regresión Logística.

En Tabla 3.1 se puede apreciar la matriz de confusión del análisis de Regresión Logística.

	Seguro	Intermedio	Crítico
Seguro	8	5	3
Intermedio	3	3	0
Crítico	3	0	2

Tabla 3. 1 Tabla de precisión del Escenario 1 de Regresión Logística.

- b) Regresión K-NN: En las Figuras 2a y 2b del Anexo 4 se puede apreciar las gráficas de la Regresión K-NN.

En Tabla 3.2 se puede apreciar la matriz de confusión del análisis de Regresión Logística.

	Seguro	Intermedio	Crítico
Seguro	13	1	2
Intermedio	0	6	0
Crítico	0	1	4

Tabla 3. 2 Tabla de precisión de la Regresión K-NN del Escenario 1.

- c) Regresión SVM: En las Figuras 3a y 3b del Anexo 4 se puede apreciar las gráficas de la Regresión SVM.

En Tabla 3.3 se puede apreciar la matriz de confusión del análisis de Regresión Logística.

	Seguro	Intermedio	Crítico
Seguro	8	5	3
Intermedio	3	3	0
Crítico	3	0	2

Tabla 3. 3 Tabla de precisión de la Regresión SVM del Escenario 1.

- d) Regresión Decision Tree: En las Figuras 4a y 4b del Anexo 4 se puede apreciar las gráficas de la Regresión SVM.

En Tabla 3.4 se puede apreciar la matriz de confusión del análisis de Regresión Logística.

	Seguro	Intermedio	Crítico
Seguro	14	0	2
Intermedio	3	3	0
Crítico	2	0	3

Tabla 3. 4 Tabla de precisión de la Regresión Decision Tree del Escenario 1.

En la tabla 3.5 se puede apreciar los ratios de precisión y error.

	Logistic	K-NN	SVM	Decision Tree
Ratio de precisión	48.15%	85.19%	48.15%	74.07%
Ratio de error	51.85%	14.81%	51.85%	25.93%

Tabla 3. 5 Ratios de precisión de cada uno de los análisis del Escenario 1.

En la tabla 3.5 se puede observar que el Análisis de Regresión K-NN es el análisis que tiene la mayor precisión, por lo tanto, es la más exacta.

3.2.2 Condición según el diámetro, tirante, ángulo, área mojada, perímetro mojado, radio hidráulico, velocidad, velocidad crítica, caudal, pendiente, número de actividades y material.

En la Tabla 2 del Anexo 3 se puede apreciar los valores tomados, solo que no se tomará en cuenta el ángulo, el área mojada, el perímetro mojado, el radio hidráulico, la velocidad crítica, pero no se tomará en cuenta la Rugosidad de Manning que es la variable correspondiente al material.

Cómo en el escenario anterior, primero se aplicará el método de Elbow para determinar la cantidad de clústeres que se va a aplicar.

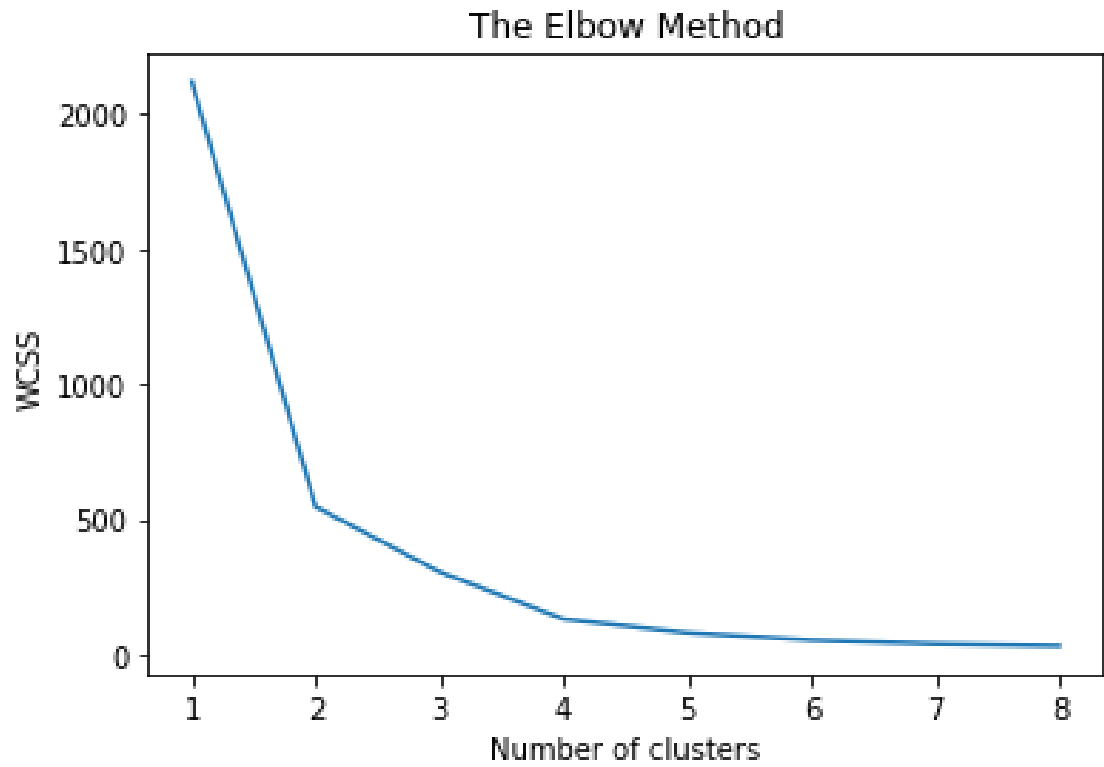


Figura 3. 5 Gráfica de método de Elbow del Escenario 2.

Según la Figura 3.5, se tomó la decisión de hacer 3 clústeres.

Viendo la Figura 3.6, se tomó la decisión de hacer 3 clústeres. En la siguiente gráfica se puede apreciar la clusterización.

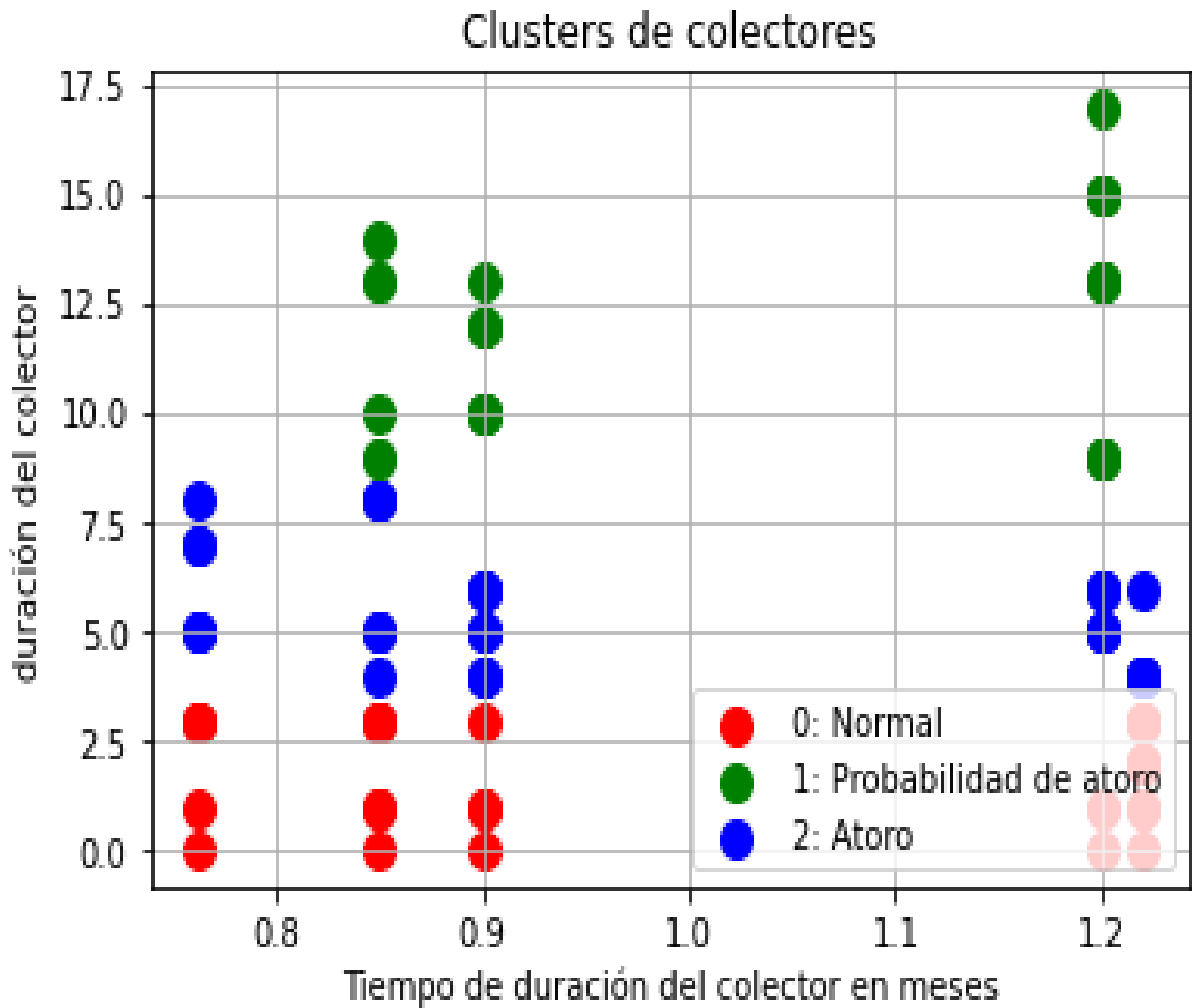


Figura 3. 6 Gráfica de Clusterización del Escenario 2.

Antes de hacer las regresiones, es necesario tener en cuenta que se debe eliminar uno de los datos, porque está constante y esa es la rugosidad de Manning, ya que el valor es constante en todo el conjunto de datos.

En la Figura 3.7 se puede apreciar el cuadro, pero de manera abreviada ya que el cuadro es demasiado grande en Excel.

	Diámetro (m)	Tiempo (meses)	Tirante de agua (m)	velocidad (m/s)	Caudal (L/s)	Pendiente (mm/mm)	Número de actividades (acumulado)	Rugosidad (n)	Manning	k-means
0	0.762	1	0.57150	4.491942	1648.008325	0.01172	0	0.009	0	
1	0.762	2	0.56388	4.483293	1622.143223	0.01172	1	0.009	0	
2	0.762	3	0.38100	3.960508	903.068640	0.01172	3	0.009	0	
3	0.762	4	0.38100	3.960508	903.068640	0.01172	3	0.009	0	
4	0.762	5	0.38100	3.960508	903.068640	0.01172	3	0.009	0	
...	
117	1.220	18	0.48800	1.506485	657.809275	0.00111	4	0.009	2	
118	1.220	19	0.48800	1.506485	657.809275	0.00111	4	0.009	2	
119	1.220	20	0.48800	1.506485	657.809275	0.00111	4	0.009	2	
120	1.220	21	0.48800	1.506485	657.809275	0.00111	4	0.009	2	
121	1.220	22	0.30500	1.168533	267.054477	0.00111	6	0.009	2	

Figura 3. 7 Resultados de la clusterización de k-mean del escenario 2.

En la Figura 3.8 se puede apreciar la correlación de toda la base de datos.

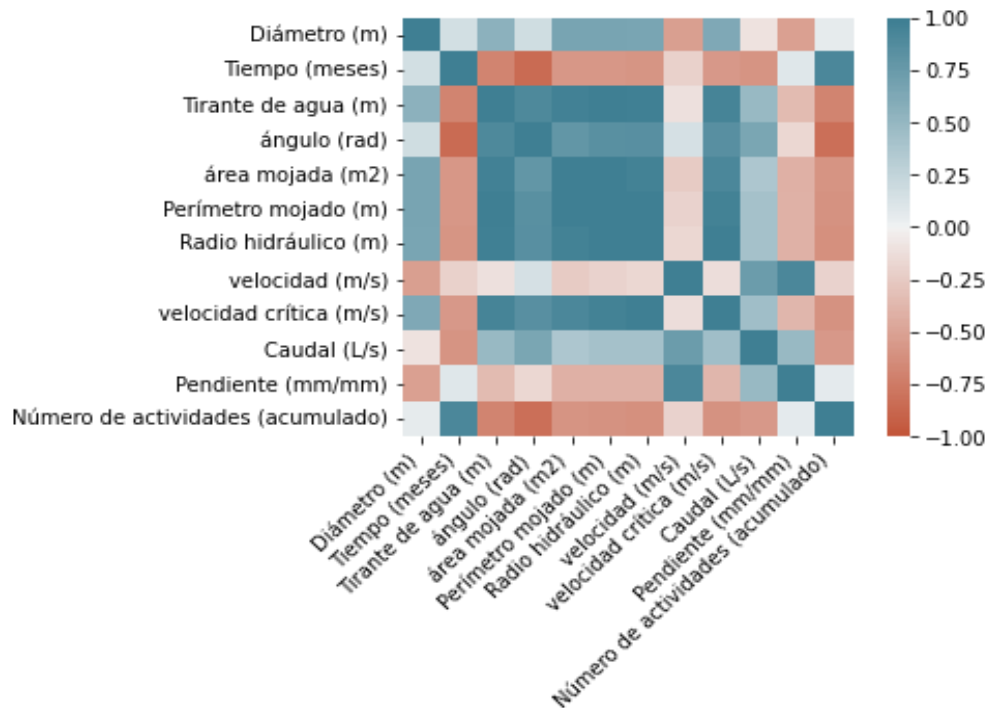


Figura 3. 8 Gráfica de Correlación del Escenario 2.

Ahora se mostrará el resultado de cada una de las regresiones tocadas en el capítulo

1.

- a) Regresión logística: En las Figura 1a y 1b del Anexo 5 se puede apreciar las gráficas de la Regresión Logística.

En Tabla 3.6 se puede apreciar la matriz de confusión del análisis de Regresión Logística.

	Normal	Probabilidad de atoro	Atoro
Normal	3	0	2
Probabilidad de atoro	0	7	4
Atoro	2	1	6

Tabla 3. 6 Matriz de Confusión del análisis de Regresión Logística del Escenario 2.

- b) Regresión K-NN: En las Figuras 2a y 2b del Anexo 5 se puede apreciar las gráficas de la Regresión Logística.

En Tabla 3.7 se puede apreciar la matriz de confusión del análisis de Regresión Logística.

	Normal	Probabilidad de atoro	Atoro
Normal	5	0	0
Probabilidad de atoro	0	10	1
Atoro	3	1	5

Tabla 3. 7 Matriz de Confusión del análisis de Regresión K-NN del Escenario 2.

- c) Regresión SVM: En las Figuras 3a y 3b del Anexo 5 se puede apreciar las gráficas de la Regresión Logística.

En la tabla 3.8 se puede apreciar la matriz de confusión del análisis de Regresión SVM.

	Normal	Probabilidad de atoro	Atoro
Normal	3	0	2
Probabilidad de atoro	0	7	4
Atoro	2	1	6

Tabla 3. 8 Matriz de Confusión del análisis de Regresión SVM del Escenario 2.

d) Regresión Decision Tree: En las Figuras 4a y 4b del Anexo 5 se puede apreciar las gráficas de la Regresión Logística.

En la tabla 3.9 se puede apreciar la matriz de confusión del análisis de Regresión SVM.

	Normal	Probabilidad de atoro	Atoro
Normal	2	0	3
Probabilidad de atoro	0	8	3
Atoro	1	0	8

Tabla 3. 9 Matriz de Confusión del análisis de Regresión Decision Tree del Escenario 2.

Una vez obtenido todas las matrices de confusión de las 4 regresiones, en la Tabla 3.10 se mostrará los ratios de cada una de las regresiones.

	Logistic	K-NN	SVM	Decision Tree
ratio de precisión	64.00%	88.00%	64.00%	72.00%
ratio de error	28.00%	12.00%	28.00%	28.00%

Tabla 3. 10 Ratios de cada uno de los análisis del Escenario 2.

Según la Tabla 3.10, el análisis de Regresión de K-NN es el más certero que los demás análisis.

CONCLUSIONES

A continuación, se presentan las conclusiones del trabajo de investigación según los resultados y también se mostrarán recomendaciones para trabajos futuros.

- Según el escenario 1: A mayor cantidad de fallas acumuladas en menos tiempo, más crítica será la condición del colector y tiene mayor riesgo de falla.
- Según el escenario 2: A mayores actividades el diámetro tiene que ser mayor.
- Se recomienda utilizar este análisis en cada uno de los distritos para evaluar su grado de criticidad de su sistema de alcantarillado y evitar problemas futuros.
- Evaluar el uso de tecnologías para la medición en los colectores con la finalidad de desarrollar nuevos métodos para el control de condición en los colectores y así generar una base de datos obtenida de estas tecnologías.
- Se recomienda con mayor cantidad de datos a fin de que los resultados se acerquen más a la realidad, pero los datos tienen que ser los adecuados y las variables deben ser evaluadas a su propio criterio.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Rosabella Palomba, “Calidad de Vida: Conceptos y Medidas, Taller sobre calidad de vida y Redes de Apoyo de las personas adultas mayores”, Institute of Population Research and Social Policies, Roma - Italia, CELADE/División de población, CEPAL, Santiago - Chile, pp, 2002.
- [2] Inei, Perú: Formas de Acceso al Agua y Saneamiento Básico, Lima- Perú, 2018.
- [3] Carlos Alberto Heredia Muñoz, “Estudio de las ineficiencias en la gestión de SEDAPAL y propuesta de una Tarifa Justa como solución. Periodo: 1996-2004”, Tesis para título profesional, Cercado de Lima, Universidad Nacional Mayor de San Marcos, Perú, 2005.
- [4] Guiuseppe Monroy Fernández de Lara, “Problemática de los Sistemas de Alcantarillado”, Tesis para especialización, Ciudad de México, Universidad Autonoma de México (UNAM), México, 2014.
- [5] Luisa Fernanda Ramirez Corredor, “Estudio y diagnóstico de la red de alcantarillado sanitario y pluvial para el proceso de densificación de un sector del centro de Bogotá”, Trabajo de grado para título profesional, Universidad Católica de Colombia, Cundinamarca, Colombia, 2016.
- [6] Narcisa de Jesús Sánchez Camino, “El modelo de Gestión y su incidencia en la provisión de los servicios de Agua potable y alcantarillado en la municipalidad de Tena”, Tesis para la obtención de Grado Académico de Magister, Universidad Técnica de Ambato, Ambato, Ecuador, 2011.
- [7] Paulo Marcos dos Anjos da Silva, “Avaliação do Sistema de Esgotamiento Sanitario da Regiao Administrativa de Sobradhino I”, Tesis de grado para título profesional, Curso de Tecnologia em Gestao Ambiental, Brasilia, 2013.

- [8] Sunil K. Sinha, Marc E. Edwards, Gerardo W. Flintsch, “Condition Assetments Technologies for Drinking Water and Wastewater Pipelines: State-of-the-Art Literature and Practice View”, Tesis para Maestría de ciencia, Virginia Polytechnic Institute and State University, Virginia, Estados Unidos, 2012.
- [9] Jean Paul Christofer Valentín Minaya, “Evaluación de sistema de Alcantarillado Pluvial con la aplicación de SewerCad en la prolongación Libertadores, Huaraz, 2018”, Tesis para título profesional, Universidad César Vallejo, Huaraz, Perú, 2018.
- [10] H. Flores T., “Informe de Emergencia N°172 - 25/02/2019 / COEN - INDECI/ 17:45 Horas”, Aniego en el Distrito de San Juan de Lurigancho - Lima”, Centro de Emergencia Nacional (COEN), Jesus Maria, Informe N°30, 2019.
- [11] Juan Hipólito Arocutipa Lorenzo, “Evaluación y propuesta técnica de una planta de tratamiento de aguas residuales en Massiapo del distrito del Alto Inambari”, Tesis para título profesional, Universidad Nacional del Altiplano – Puno, Perú, 2013.
- [12] Chirinos Alvarado, Shirly Bibi, “Diseño del Sistema de abastecimiento de agua potable y alcantarillado del Caserío Anta, Moro - Ancash 2017”, Tesis para título profesional, Universidad César Vallejo (UCV), Chimbote, Perú, 2017.
- [13] Christian Israel Vicario Flores, David Herrera Domínguez, Guillermo Gaitán Diéguez, Mario Alberto Zúñica Cortés, “Propuesta de una red industrial de monitoreo para facilidades eléctricas”, tesis para título profesional, Universidad Nacional Autónoma de México, Ciudad de México, México, 2015.
- [14] Rafael Delgadillo Porras, “Proyecto de alcantarillado de aguas residuales de la localidad de Quiahuitepec, Municipio de Ayutla de los Libres, Estado Guerrero”, Tesis para título profesional, Instituto Politécnico Nacional, México, 2009.
- [15] Sedapal, “OBRAS GENERALES PARA LA INSTALACIÓN DEL COLECTOR DE ALIVIO PARA EL COLECTOR N°6 DEL DISTRITO DEL RIMAC”, nov. 2017

- [16] “Procedimiento Constructivo: Para la actualización del estudio definitivo y expediente técnico del proyecto: Obras Generales Para la Instalación del Colector de Alivio Para el Colector N°6 – Distritos de Rímac y San Martín de Porres”, SEDAPAL, Perú, 2017.
- [17] RBF Consulting, “Sanitary Sewer Maintenance Plan, City of Imperial Beach”, Imperial Beach CA, California, Estados Unidos, 2014.
- [18] Aldo Alex Paniagua Carty y Harold Edu Quispe Serpa, “Estudio Comparativo entre el método de excavación sin zanja y el método de excavación tradicional para el cambio de tuberías de agua potable y desagüe”, Tesis para título profesional, Universidad San Martín de Porres (USMP), Lima, Perú, 2017.
- [19] José Luis Mata Romualdo, “Investigación experimental de flujo transitorio mixto debido al llenado rápido de un sistema de alcantarillado que contiene aire atrapado”, tesis para título profesional, Universidad Nacional Autónoma de México (UNAM), Ciudad de México, México, 2014.
- [20] INEI, “Lima alberga 9 millones 320 mil habitantes”, en. 2018 [online]. Disponible: <https://www.inei.gob.pe/prensa/noticias/lima-alberga-9-millones-320-mil-habitantes-al-2018-10521/>
- [21] “Plan de Operaciones de emergencia de San Juan de Lurigancho – POE, 2017-2018”, Municipalidad de San Juan de Lurigancho, Lima, Perú, 2017.
- [22] Comisión del Gobierno Peruano, Contigo Ciudadano, “La Ciudad de Lima en Cifras”, 2018 [offline].
- [23] “Memoria descriptiva para la actualización del estudio definitivo y expediente técnico del proyecto: Obras generales para la instalación del colector de alivio para el colector N°06”, SEDAPAL, Lima, Perú, 2017.

- [24] Medrano Prado Huarcaya, “Manual de operación y mantenimiento (colectores y conexiones domiciliarias), sistema de alcantarillado”, SEDAPAL, El Agustino, Manual, 2016.
- [25] Juan Manuel Escallón Arango, “Determinación del estado de redes al alcantarillado teniendo en cuenta inspecciones con circuito de televisión (CCTV)”, Proyecto de grado para título profesional, Universidad de los Andes, Bogotá, Colombia, 2005.
- [26] Sonia Pighin, “Informática Aplicada a la Ingeniería de Procesos I (Orientación I)”, Universidad Tecnológica Nacional, Buenos Aires, Argentina, 2001.
- [27] Ana del Pozo Santolaya, “Herramientas de inteligencia de negocio”, trabajo de fin de grado, Universidad de la Roja, España, 2017.
- [28] Mohseen Mohammed, Muhammad Badruddin Khan, Eihab Bashier Mohammed Bashier, “Machine Learning Algorithms Applications”, 2017.
- [29] SEDAPAL, “Formato N°5 Producción de Agua Potable Al 31 de diciembre de 2018”, dic. 2018.
- [30] SEDAPAL, “Formato N°3 Tratamiento de Aguas Servidas al 31.12.2018”, en. 2019.
- [31] Andie Bryan Dongo Román, “Descripción Metodológica del análisis clúster utilizando el algoritmo Ward”, trabajo monográfico para optar por el título de ingeniero estadístico e informático, Universidad Nacional Agraria La Molina, Lima, Perú, 2017.
- [32] Henry Lamos Díaz, “Aplicación de técnicas de agrupamiento (clustering) para el análisis estadístico de tendencias en twitter basado en el lenguaje de programación R”, trabajo de grado para optar el título de Ingeniero Industrial, Universidad Industrial Santander, Bucaramanga, Colombia, 2017.
- [33] Juan Zamorano Ruiz, “Comparativa y Análisis de algoritmos de aprendizaje automático para la predicción del tipo predominante de cubierta arbórea”, trabajo Fin Master en

Ingeniería de Computadores, Universidad Complutense de Madrid, Madrid, España 2018.

- [34] Christopher Alejandro Flores Jara, “Extracción de información en EMR para la Identificación de Obesidad mediante el Estudio de Comorbilidades Asociadas”, Informe de Memoria de Título para optar al título de Ingeniero Civil Biomédico, Universidad de Concepción, Perú, 2015.
- [35] Bach. Madueño Auris Alcides Aníbal, “Diseño y ejecución del canal-Túnel para la evacuación de aguas pluviales en Cutervo - Cajamarca”, trabajo monográfico para optar el título de Ingeniero Agrícola, Universidad Nacional Agraria La Molina, Lima, Perú, 2018.
- [36] Joseph Baldir Zuñiga Ancasi, “Verificación hidráulica – Aplicación del Sistema ISO 14001 y programación en ritmo constante para la obra: Ampliación y mejoramiento de los sistemas de agua potable y alcantarillado del sector El Triunfo que comprende Ocho asentamientos humanos – distrito La Joya, provincia y región de Arequipa”, tesis para título profesional, Universidad San Agustín, Arequipa, Perú, 2017.
- [37] Guilherme Firmino de Carvalho Nunes de Carvalho, “Data Mining techniques to predict sewer condition”, Tesis para maestría de Ciencias, Técnico Lisboa, Portugal, 2015.
- [38] Orlando Belli Hesse, “Modelos predictivos de Machine Learning con R y/o Python”, ESCUELA DE FORMACIÓN PROFESIONAL Ingenium, oct. 2020, clase 7.
- [39] Orlando Belli Hesse, ESCUELA DE FORMACIÓN PROFESIONAL Ingenium, oct. 2020, clase 9 y Clase 10.
- [40] Información brindada por el área de investigación de SEDAPAL
- [41] SEDAPAL, Plano brindado por la empresa.

ANEXOS

ANEXO 1: PRODUCCIÓN DE AGUA POTABLE Y TRATAMIENTO DE AGUAS SERVIDAS DEL 2018

Localidad	Tipo de Planta	Producción en m3
El Agustino - Lima - Lima	Planta N°1 – Superficial	246575
El Agustino - Lima - Lima	Planta N°2 – Superficial	276842
Carabayllo - Lima - Lima	Planta chillón – Superficial	25174
Ate Vitarte - Lima - Lima	Pozos Huachipa – Superficial	36958
Centro de servicios Comas	Pozos SEDAPAL – Subterránea	14486
Centro de servicios Callao	Pozos SEDAPAL – Subterránea	17095
Centro de servicios Ate Vitarte	Pozos SEDAPAL – Subterránea	50239
Centro de servicios Breña	Pozos SEDAPAL – Subterránea	2942
Centro de servicios San Juan de Lurigancho	Pozos SEDAPAL – Subterránea	20371
Centro de servicios Surquillo	Pozos SEDAPAL – Subterránea	9334
Centro de servicios Villa El Salvador	Pozos SEDAPAL – Subterránea	9795
Centro de servicios Comas	Pozos Chillón – Subterránea	7652
Centro de servicios Callao	Pozos Chillón – Subterránea	11863
Total		729326

Tabla 1 Producción de agua potable del 2018 [29].

Localidad	Porcentaje de Tratamiento (*)
Lima - Callao	92.20%

Tabla 2 Porcentaje de tratamiento de agua residual del 2018 [30].

ANEXO 2: TABLA DE ÍNDICE DE INCIDENCIAS EN SJL SEGÚN SEDAPAL

Fuente “Carta N°333-2019-ESG, Atención a Solicitud de Acceso a la Información Pública”, SEDAPAL, solicitado 20 de junio del 2019, ERprim, tomado el 7 de julio del 2019

Incidencia	Distrito	Fecha	Ubicación	Observaciones	Motivo	Tramo/Bz	Colector	DN(mm)	L(m)	SED.	TIR.	Sub.Act.
EVALUAR LA CAPACIDAD DE LOS COLECTORES PRIMARIOS	S.J. DE LURIGANCHO	12-sep-18	mariscanías / nectandros	Se hizo levantamiento topográfico	OTRO	Bz-264283	Emisor Sur	0	0	0	0	Recolectar datos topográficos
REHABILITACIÓN DE COLECTORES PRIMARIOS	S.J. DE LURIGANCHO	19-ene-17	av. proceres de la independencia ctra. 10		COLAPSO DE TUBERIA	GGD-132856	Huayrona	600	0	0	0	Bombeo de aguas residuales, Robura y excavación de zanja
REHABILITACIÓN DE COLECTORES PRIMARIOS	S.J. DE LURIGANCHO	20-ene-17	av. proceres de la independencia ctra. 10		COLAPSO DE TUBERIA	GGD-132856	Huayrona	600	24	0	0	Instalación de tuberías sin zanja
REHABILITACIÓN DE COLECTORES PRIMARIOS	S.J. DE LURIGANCHO	20-ene-17	av. proceres de la independencia ctra. 10		COLAPSO DE TUBERIA	GGD-132856	Huayrona	600	0	0	0	Instalación de tuberías
REHABILITACIÓN DE COLECTORES PRIMARIOS	S.J. DE LURIGANCHO	20-ene-17	av. proceres de la independencia ctra. 10		COLAPSO DE TUBERIA	GGD-132856	Huayrona	600	24	0	0	Instalación de tuberías sin zanja
REHABILITACIÓN DE COLECTORES PRIMARIOS	S.J. DE LURIGANCHO	20-ene-17	av. proceres de la independencia ctra. 10		COLAPSO DE TUBERIA	GGD-132856	Huayrona	600	24	0	0	Bombeo de aguas residuales, Instalación de tuberías, Relleno y compactación
REHABILITACIÓN DE COLECTORES PRIMARIOS	S.J. DE LURIGANCHO	23-ene-17	av. proceres de la independencia ctra. 13	TAMBIEN RELLENO DE ZANJA EN LA CDRA. 9	COLAPSO DE TUBERIA	GGD-132865	Huayrona	600	0	0	0	Robura y excavación de zanja
EJECUTAR PIQUES Y/O DESATOS	S.J. DE LURIGANCHO	31-oct-17	av. los posas esp. av. proceres de la independencia	El centro de servicios S.J. de Lurigancho, aviso a las 08:30am. Al llegar, ya estuvo solucionado.	PIQUE Y DESATORO	Bz-24099	La Huayrona	0	0	0	0	Desatoro con hidrojete, Desatoro con vanilla
MANTENIMIENTO DE BUZONES	S.J. DE LURIGANCHO	19-ene-17	av. proceres de la independencia ctra. 12	SELLADO DE TUBERIA DENTRO DEL BUZON PARA INSTALACION DE LA NUEVA TUBERIA	COLAPSO DE TUBERIA	Bz-261561	Huayrona	0	0	0	0	Limpieza manual de buzon, Sellado de tuberia
MANTENIMIENTO DE BUZONES	S.J. DE LURIGANCHO	08-jul-18	av. proceres de la independencia 1057		OTRO	Bz-261047	Canto Grande	0	0	0	0	Mantenimiento de tapa de buzon
CAMBIO DE MARCO Y/O TAPA EN COLECTORES PRIMARIOS	S.J. DE LURIGANCHO	05-feb-18	av. proceres de la independencia esp. BAZONER	Tambien realizaron trabajos en base.	OTRO	Bz-251948	Nuevo Bayovar	0	0	0	0	Rotura de pavimento para Cambio de marco y de tapa, Eliminación de elemento. Resque de pavimento.

ANEXO 3: Cuadros de los escenarios 1 y 2

Diámetro (mm)	longitud (m)	Pendiente	Profundidad (m)	Tiempo (años)	Número de fallas (acumulativo)	Número de fallas
762	1000	0.01172	9	1	1	1
762	1000	0.01172	9	2	3	2
762	1000	0.01172	9	3	3	0
762	1000	0.01172	9	4	3	0
762	1000	0.01172	9	5	4	1
762	1000	0.01172	9	6	4	0
762	1000	0.01172	9	7	4	0
762	1000	0.01172	9	8	6	2
762	1000	0.01172	9	9	6	0
762	1000	0.01172	9	10	6	0
762	1000	0.01172	9	11	7	1
762	1000	0.01172	9	12	7	0
762	1000	0.01172	9	13	7	0
762	1000	0.01172	9	14	8	1
762	1000	0.01172	9	15	8	0
762	1000	0.01172	9	16	8	0
762	1000	0.01172	9	17	8	0
762	1000	0.01172	9	18	8	0
762	1000	0.01172	9	19	8	0
762	1000	0.01172	9	20	8	0
762	1000	0.01172	9	21	9	1
762	1000	0.01172	9	22	9	0
762	1000	0.01172	9	23	9	0
762	1000	0.01172	9	24	9	0
762	1000	0.01172	9	25	9	0
762	1000	0.01172	9	26	9	0
762	1000	0.01172	9	27	9	0
762	1000	0.01172	9	28	9	0
762	1000	0.01172	9	29	10	1
762	1000	0.01172	9	30	10	0
762	1000	0.01172	9	31	10	0
762	1000	0.01172	9	32	10	0
762	1000	0.01172	9	33	10	0
762	1000	0.01172	9	34	10	0
762	1000	0.01172	9	35	11	1
762	1000	0.01172	9	36	11	0

762	1000	0.01172	9	37	11	0
762	1000	0.01172	9	38	12	1
762	1000	0.01172	9	39	12	0
762	1000	0.01172	9	40	12	0
762	16365	0.01172	10	1	0	0
850	16365	0.00656	10	2	1	1
850	16365	0.00656	10	3	1	0
850	16365	0.00656	10	4	3	2
850	16365	0.00656	10	5	6	3
850	16365	0.00656	10	6	6	0
850	16365	0.00656	10	7	7	1
850	16365	0.00656	10	8	9	2
850	16365	0.00656	10	9	10	1
850	16365	0.00656	10	10	13	3
850	16365	0.00656	10	11	17	4
850	16365	0.00656	10	12	18	1
850	16365	0.00656	10	13	18	0
850	16365	0.00656	10	14	18	0
850	16365	0.00656	10	15	19	1
850	16365	0.00656	10	16	20	1
850	16365	0.00656	10	17	21	1
850	16365	0.00656	10	18	21	0
850	16365	0.00656	10	19	22	1
850	16365	0.00656	10	20	24	2
850	16365	0.00656	10	22	26	2
850	16365	0.00656	10	23	27	1
900	16484.65	0.04656	12	1	0	0
900	16484.65	0.04656	12	2	1	1
900	16484.65	0.04656	12	3	3	2
900	16484.65	0.04656	12	4	4	1
900	16484.65	0.04656	12	5	6	2
900	16484.65	0.04656	12	6	8	2
900	16484.65	0.04656	12	7	11	3
900	16484.65	0.04656	12	8	12	1
900	16484.65	0.04656	12	9	16	4
900	16484.65	0.04656	12	10	17	1
900	16484.65	0.04656	12	11	18	1
900	16484.65	0.04656	12	12	20	2
900	16484.65	0.04656	12	13	21	1
900	16484.65	0.04656	12	14	22	1

900	16484.65	0.04656	12	15	23	1
900	16484.65	0.04656	12	16	24	1
900	16484.65	0.04656	12	17	25	1
900	16484.65	0.04656	12	18	26	1
900	16484.65	0.04656	12	19	27	1
900	16484.65	0.04656	12	20	28	1
900	16484.65	0.04656	12	21	29	1
1200	17309.04	0.00134	13	1	0	0
1200	17309.04	0.00134	13	2	0	0
1200	17309.04	0.00134	13	3	1	1
1200	17309.04	0.00134	13	4	1	0
1200	17309.04	0.00134	13	5	2	1
1200	17309.04	0.00134	13	6	2	0
1200	17309.04	0.00134	13	7	2	0
1200	17309.04	0.00134	13	8	2	0
1200	17309.04	0.00134	13	9	2	0
1200	17309.04	0.00134	13	10	2	0
1200	17309.04	0.00134	13	11	2	0
1200	17309.04	0.00134	13	12	4	2
1200	17309.04	0.00134	13	13	5	1
1200	17309.04	0.00134	13	14	6	1
1200	17309.04	0.00134	13	15	7	1
1200	17309.04	0.00134	13	16	8	1
1200	17309.04	0.00134	13	17	10	2
1200	17309.04	0.00134	13	18	10	0
1200	17309.04	0.00134	13	19	10	0
1200	17309.04	0.00134	13	20	11	1
1200	17309.04	0.00134	13	21	12	1
1200	17309.04	0.00134	13	22	12	0
1200	17309.04	0.00134	13	23	12	0
1200	17309.04	0.00134	13	24	12	0
1200	17309.04	0.00134	13	25	12	0
1200	17309.04	0.00134	13	26	13	1
1200	17309.04	0.00134	13	27	13	0
1200	17309.04	0.00134	13	28	14	1
1200	17309.04	0.00134	13	29	15	1
1200	17309.04	0.00134	13	30	15	0
1200	17309.04	0.00134	13	31	15	0
1200	17309.04	0.00134	13	32	15	0
1200	17309.04	0.00134	13	33	15	0

1200	17309.04	0.00134	13	34	15	0
1200	17309.04	0.00134	13	35	15	0
1220	1800	0.00111	14	1	1	1
1220	1800	0.00111	14	2	1	0
1220	1800	0.00111	14	3	3	2
1220	1800	0.00111	14	4	3	0
1220	1800	0.00111	14	5	6	3
1220	1800	0.00111	14	6	9	3
1220	1800	0.00111	14	7	9	0
1220	1800	0.00111	14	8	10	1
1220	1800	0.00111	14	9	12	2
1220	1800	0.00111	14	10	13	1
1220	1800	0.00111	14	11	13	0
1220	1800	0.00111	14	12	13	0
1220	1800	0.00111	14	13	13	0
1220	1800	0.00111	14	14	14	1
1220	1800	0.00111	14	15	14	0
1220	1800	0.00111	14	16	15	1

Tabla 1 Valores asumidos para cada variable para el Escenario 1 [37] [40] [41] [Elaboración propia]

Diámetro (m)	Tiempo (meses)	Tirante de agua (m)	velocidad (m/s)	Caudal (l/s)	Pendiente (mm/mm)	Número de actividades (acumulado)	Rugosidad (n) Manning
0.762	1	0.5715	4.49194187	1648.00833	0.01172	0	0.009
0.762	2	0.56388	4.48329301	1622.14322	0.01172	1	0.009
0.762	3	0.381	3.96050835	903.06864	0.01172	3	0.009
0.762	4	0.381	3.96050835	903.06864	0.01172	3	0.009
0.762	5	0.381	3.96050835	903.06864	0.01172	3	0.009
0.762	6	0.381	3.96050835	903.06864	0.01172	3	0.009
0.762	7	0.381	3.96050835	903.06864	0.01172	3	0.009
0.762	8	0.29718	3.52641154	580.682048	0.01172	5	0.009
0.762	9	0.29718	3.52641154	580.682048	0.01172	5	0.009
0.762	10	0.29718	3.52641154	580.682048	0.01172	5	0.009
0.762	11	0.29718	3.52641154	580.682048	0.01172	5	0.009
0.762	12	0.1524	2.4300377	157.781913	0.01172	7	0.009
0.762	13	0.1524	2.4300377	157.781913	0.01172	7	0.009
0.762	14	0.1524	2.4300377	157.781913	0.01172	7	0.009
0.762	15	0.0381	1.0105348	8.61451747	0.01172	8	0.009
0.85	1	0.63750	3.61595562	1650.73087	0.00656	0	0.009
0.85	2	0.57800	3.5483082	1458.0307	0.00656	1	0.009

0.85	3	0.57800	3.5483082	1458.0307	0.00656	1	0.009
0.85	4	0.57800	3.5483082	1458.0307	0.00656	1	0.009
0.85	5	0.55250	3.50640571	1369.08234	0.00656	3	0.009
0.85	6	0.55250	3.50640571	1369.08234	0.00656	3	0.009
0.85	7	0.55250	3.50640571	1369.08234	0.00656	3	0.009
0.85	8	0.55250	3.50640571	1369.08234	0.00656	3	0.009
0.85	9	0.50150	3.40061136	1184.76356	0.00656	3	0.009
0.85	10	0.48450	3.35895805	1122.34338	0.00656	4	0.009
0.85	11	0.48450	3.35895805	1122.34338	0.00656	5	0.009
0.85	12	0.48450	3.35895805	1122.34338	0.00656	5	0.009
0.85	13	0.38250	3.04197429	753.376526	0.00656	5	0.009
0.85	14	0.38250	3.04197429	753.376526	0.00656	8	0.009
0.85	15	0.38250	3.04197429	753.376526	0.00656	8	0.009
0.85	16	0.38250	3.04197429	753.376526	0.00656	8	0.009
0.85	17	0.25500	2.4713088	353.833685	0.00656	8	0.009
0.85	18	0.25500	2.4713088	353.833685	0.00656	9	0.009
0.85	19	0.19550	2.12461332	209.478337	0.00656	9	0.009
0.85	20	0.16150	1.89691198	142.397086	0.00656	10	0.009
0.85	21	0.16150	1.89691198	142.397086	0.00656	13	0.009
0.85	22	0.16150	1.89691198	142.397086	0.00656	13	0.009
0.85	23	0.04250	0.81346755	8.62874884	0.00656	14	0.009
0.9	1	0.675	10.0094326	5122.82839	0.04656	0	0.009
0.9	2	0.666	9.99016024	5042.42681	0.04656	1	0.009
0.9	3	0.666	9.99016024	5042.42681	0.04656	1	0.009
0.9	4	0.666	9.99016024	5042.42681	0.04656	1	0.009
0.9	5	0.549	9.51983417	3869.4528	0.04656	3	0.009
0.9	6	0.522	9.3567787	3579.98498	0.04656	4	0.009
0.9	7	0.522	9.3567787	3579.98498	0.04656	4	0.009
0.9	8	0.522	9.3567787	3579.98498	0.04656	4	0.009
0.9	9	0.522	9.3567787	3579.98498	0.04656	4	0.009
0.9	10	0.441	8.74883034	2712.02189	0.04656	5	0.009
0.9	11	0.441	8.74883034	2712.02189	0.04656	5	0.009
0.9	12	0.441	8.74883034	2712.02189	0.04656	5	0.009
0.9	13	0.441	8.74883034	2712.02189	0.04656	5	0.009
0.9	14	0.36	7.95776863	1891.00093	0.04656	6	0.009
0.9	15	0.36	7.95776863	1891.00093	0.04656	6	0.009
0.9	16	0.36	7.95776863	1891.00093	0.04656	6	0.009
0.9	17	0.36	7.95776863	1891.00093	0.04656	6	0.009
0.9	18	0.36	7.95776863	1891.00093	0.04656	6	0.009
0.9	19	0.261	6.71367015	1028.05527	0.04656	10	0.009

0.9	20	0.261	6.71367015	1028.05527	0.04656	10	0.009
0.9	21	0.261	6.71367015	1028.05527	0.04656	10	0.009
0.9	22	0.261	6.71367015	1028.05527	0.04656	10	0.009
0.9	23	0.261	6.71367015	1028.05527	0.04656	10	0.009
0.9	24	0.261	6.71367015	1028.05527	0.04656	10	0.009
0.9	25	0.261	6.71367015	1028.05527	0.04656	10	0.009
0.9	26	0.135	4.54576229	272.011666	0.04656	12	0.009
0.9	27	0.135	4.54576229	272.011666	0.04656	12	0.009
0.9	28	0.135	4.54576229	272.011666	0.04656	12	0.009
0.9	29	0.135	4.54576229	272.011666	0.04656	12	0.009
0.9	30	0.099	3.74391836	142.548022	0.04656	13	0.009
1.2	1	0.9	2.05903783	1873.45018	0.00134	0	0.009
1.2	2	0.876	2.05058047	1813.95235	0.00134	1	0.009
1.2	3	0.876	2.05058047	1813.95235	0.00134	1	0.009
1.2	4	0.792	2.00508229	1587.80821	0.00134	5	0.009
1.2	5	0.792	2.00508229	1587.80821	0.00134	5	0.009
1.2	6	0.792	2.00508229	1587.80821	0.00134	5	0.009
1.2	7	0.792	2.00508229	1587.80821	0.00134	5	0.009
1.2	8	0.792	2.00508229	1587.80821	0.00134	5	0.009
1.2	9	0.792	2.00508229	1587.80821	0.00134	5	0.009
1.2	10	0.768	1.98776465	1519.48816	0.00134	5	0.009
1.2	11	0.636	1.85984029	1132.01177	0.00134	6	0.009
1.2	12	0.636	1.85984029	1132.01177	0.00134	6	0.009
1.2	13	0.636	1.85984029	1132.01177	0.00134	6	0.009
1.2	14	0.636	1.85984029	1132.01177	0.00134	6	0.009
1.2	15	0.528	1.71413062	821.573356	0.00134	6	0.009
1.2	16	0.528	1.71413062	821.573356	0.00134	9	0.009
1.2	17	0.528	1.71413062	821.573356	0.00134	9	0.009
1.2	18	0.528	1.71413062	821.573356	0.00134	9	0.009
1.2	19	0.528	1.71413062	821.573356	0.00134	9	0.009
1.2	20	0.528	1.71413062	821.573356	0.00134	9	0.009
1.2	21	0.528	1.71413062	821.573356	0.00134	9	0.009
1.2	22	0.42	1.52906168	539.409977	0.00134	13	0.009
1.2	23	0.42	1.52906168	539.409977	0.00134	13	0.009
1.2	24	0.42	1.52906168	539.409977	0.00134	13	0.009
1.2	25	0.42	1.52906168	539.409977	0.00134	13	0.009
1.2	26	0.42	1.52906168	539.409977	0.00134	13	0.009
1.2	27	0.3	1.26976208	280.752707	0.00134	15	0.009
1.2	28	0.3	1.26976208	280.752707	0.00134	15	0.009
1.2	29	0.3	1.26976208	280.752707	0.00134	15	0.009

1.2	30	0.3	1.26976208	280.752707	0.00134	15	0.009
1.2	31	0.3	1.26976208	280.752707	0.00134	15	0.009
1.2	32	0.24	1.11389232	179.365934	0.00134	17	0.009
1.22	1	0.915	1.89488591	1782.04251	0.00111	0	0.009
1.22	2	0.8906	1.8871028	1725.44764	0.00111	1	0.009
1.22	3	0.8906	1.8871028	1725.44764	0.00111	1	0.009
1.22	4	0.8906	1.8871028	1725.44764	0.00111	1	0.009
1.22	5	0.793	1.83747797	1477.98952	0.00111	2	0.009
1.22	6	0.793	1.83747797	1477.98952	0.00111	2	0.009
1.22	7	0.793	1.83747797	1477.98952	0.00111	2	0.009
1.22	8	0.793	1.83747797	1477.98952	0.00111	2	0.009
1.22	9	0.793	1.83747797	1477.98952	0.00111	2	0.009
1.22	10	0.6954	1.76021029	1211.62308	0.00111	3	0.009
1.22	11	0.6954	1.76021029	1211.62308	0.00111	3	0.009
1.22	12	0.6954	1.76021029	1211.62308	0.00111	3	0.009
1.22	13	0.6954	1.76021029	1211.62308	0.00111	3	0.009
1.22	14	0.488	1.50648536	657.809275	0.00111	4	0.009
1.22	15	0.488	1.50648536	657.809275	0.00111	4	0.009
1.22	16	0.488	1.50648536	657.809275	0.00111	4	0.009
1.22	17	0.488	1.50648536	657.809275	0.00111	4	0.009
1.22	18	0.488	1.50648536	657.809275	0.00111	4	0.009
1.22	19	0.488	1.50648536	657.809275	0.00111	4	0.009
1.22	20	0.488	1.50648536	657.809275	0.00111	4	0.009
1.22	21	0.488	1.50648536	657.809275	0.00111	4	0.009
1.22	22	0.305	1.1685333	267.054477	0.00111	6	0.009

Tabla 2 Tabla de datos para el Escenario 2 [35] [36] [40] [41] [Elaboración propia]

ANEXO 4: Gráficas de cada una de las regresiones del escenario 1

a) Regresión Logística:

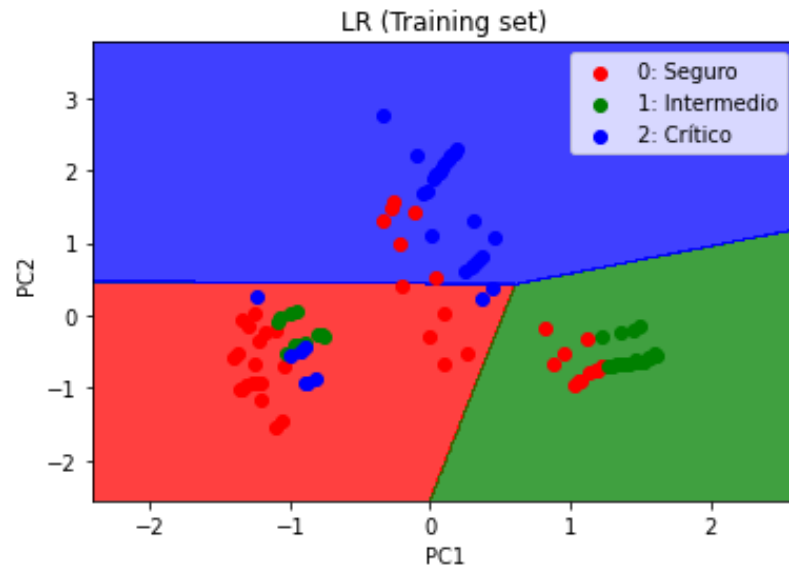


Figura 1a Gráfica del análisis de Regresión Logística (training set) del escenario 1.

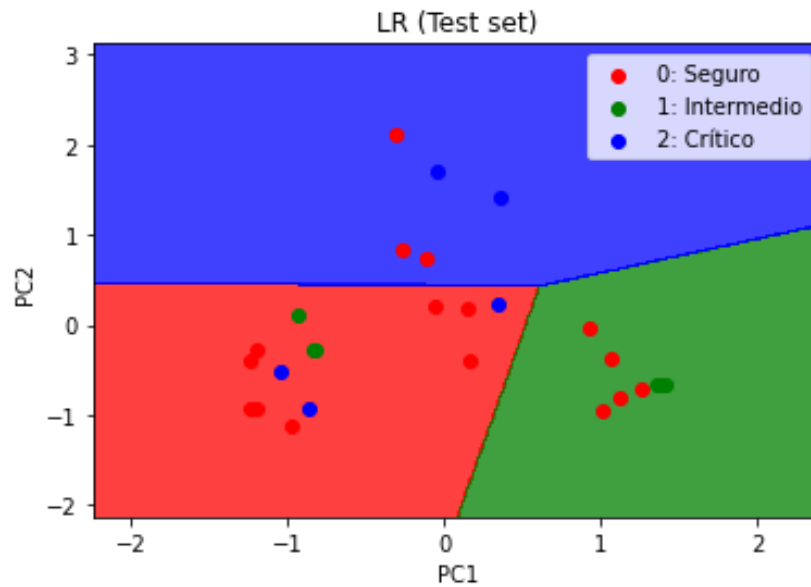


Figura 1b Gráfica del análisis de Regresión Logística (Test set) del escenario 1.

b) Regresión K-NN:

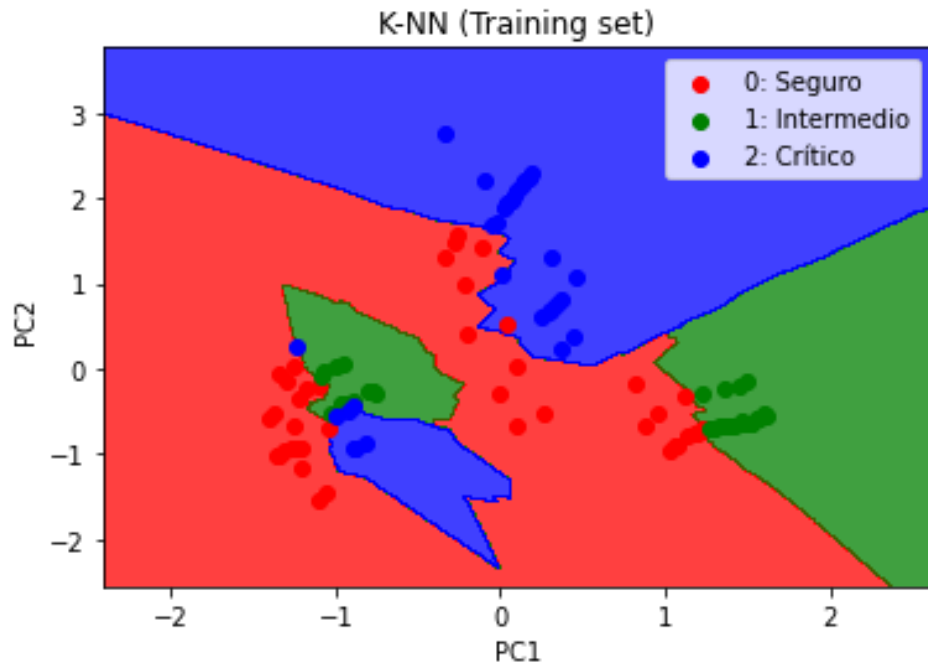


Figura 2a Gráfica del análisis de Regresión K-NN (Training set) del escenario 1.

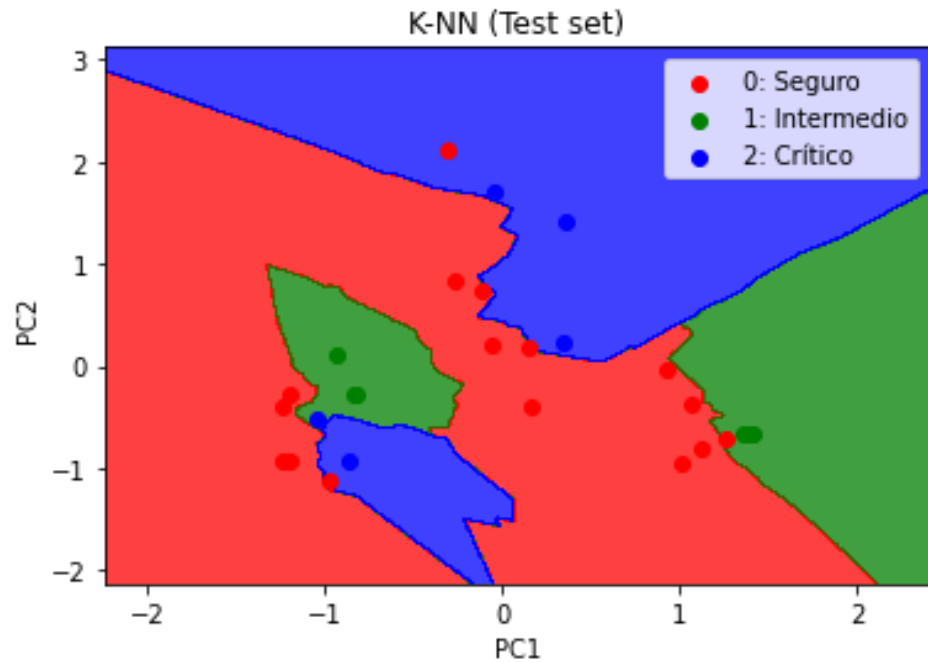


Figura 2b Gráfica del análisis de Regresión K-NN (Test set) del escenario 1.

c) Regresión SVM:

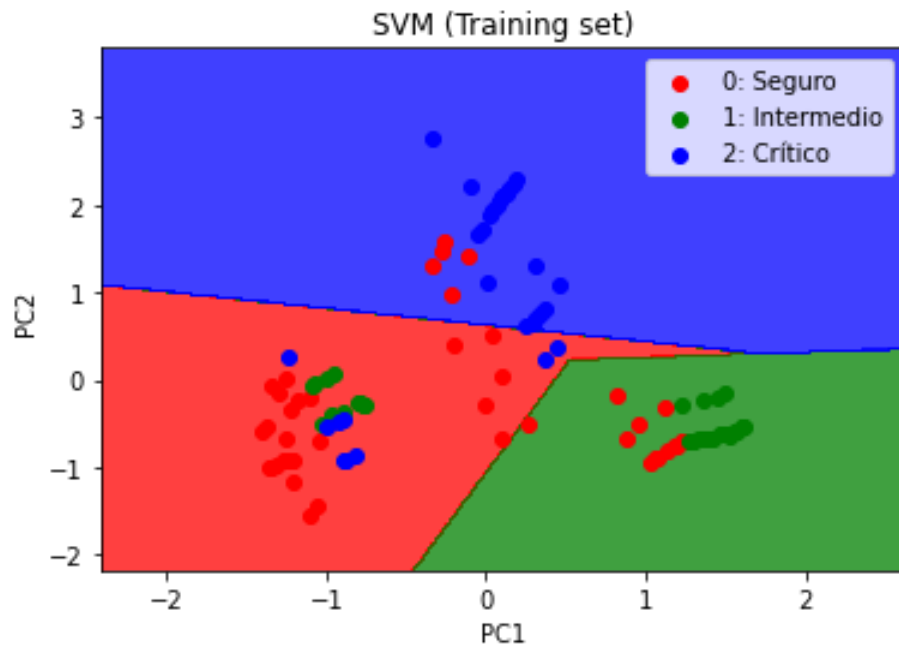


Figura 3a Gráfica del análisis de Regresión SVM (Training set) del escenario 1.

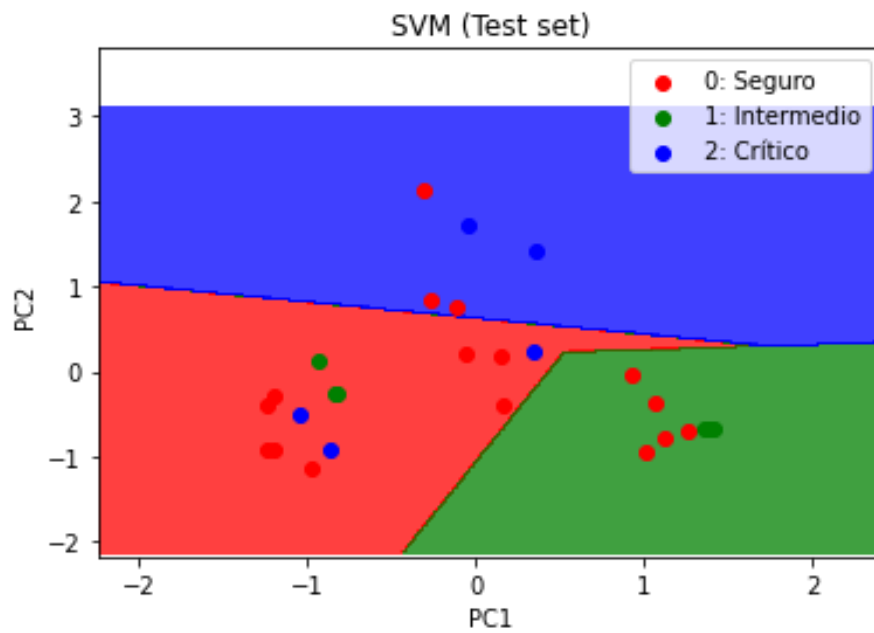


Figura 3b Gráfica del análisis de Regresión SVM (Test set) del escenario 1.

d) Regresión Decision Tree:

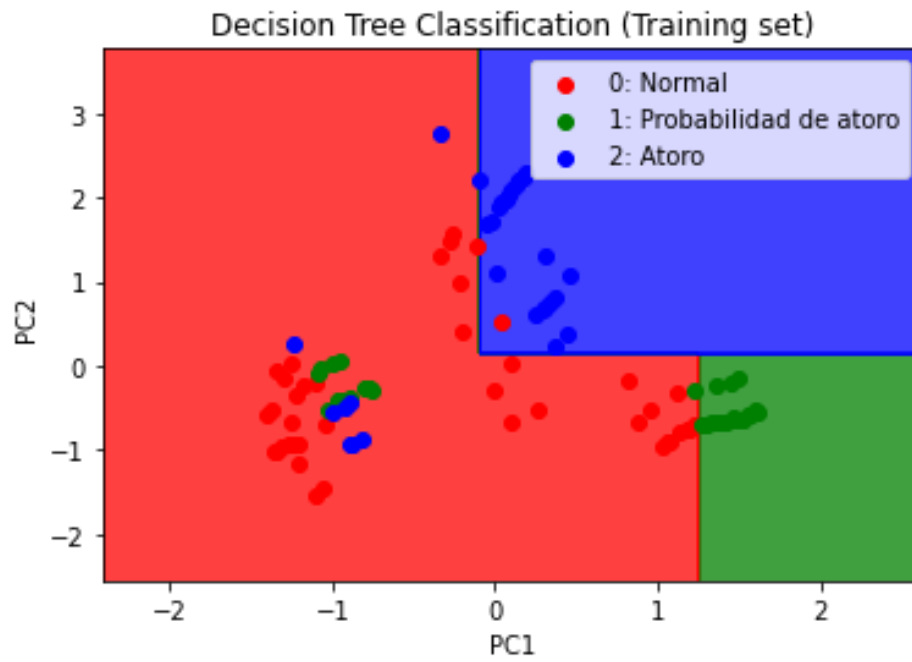


Figura 4a Gráfica del análisis de Regresión Decision Tree (Training set) del escenario 1.

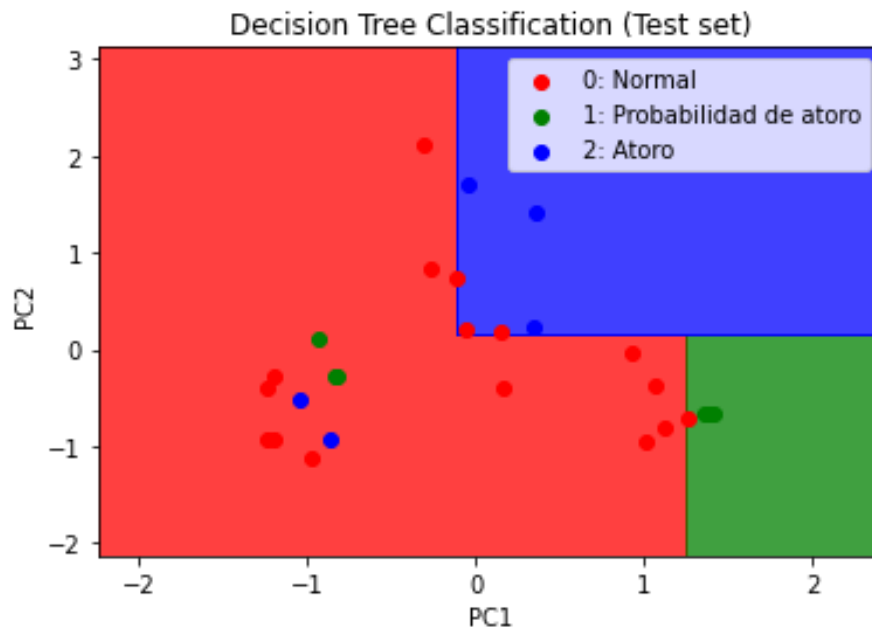


Figura 4a Gráfica del análisis de Regresión Decision Tree (Test set) del escenario 1.

ANEXO 5: Gráficas de cada una de los regresiones del Escenario 2

a) Regresión Logística:

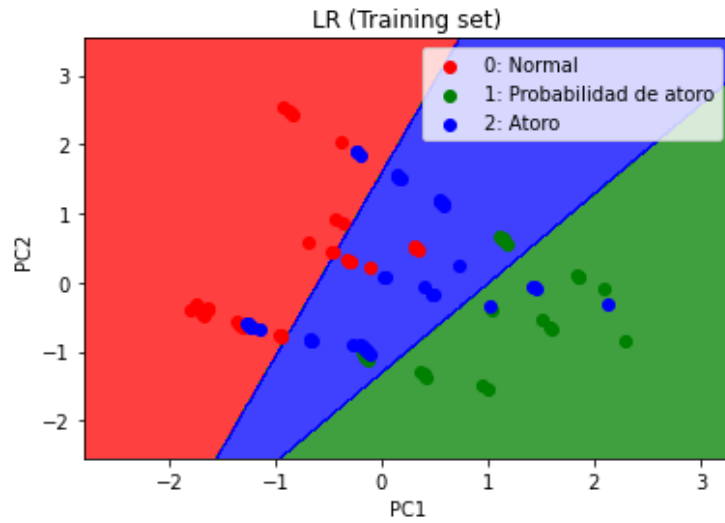


Figura 1a Gráfica del análisis de Regresión Logística (Training set) del escenario 2.

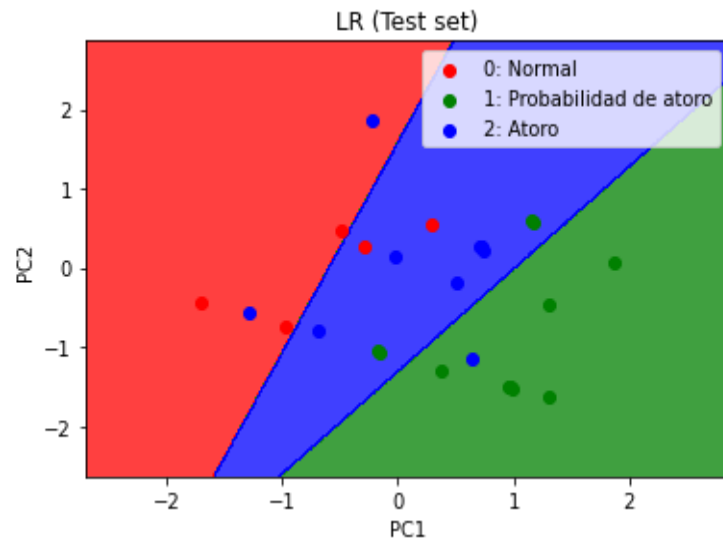


Figura 1b Gráfica del análisis de Regresión Logística (Test set) del escenario 2.

b) Regresión K-NN:

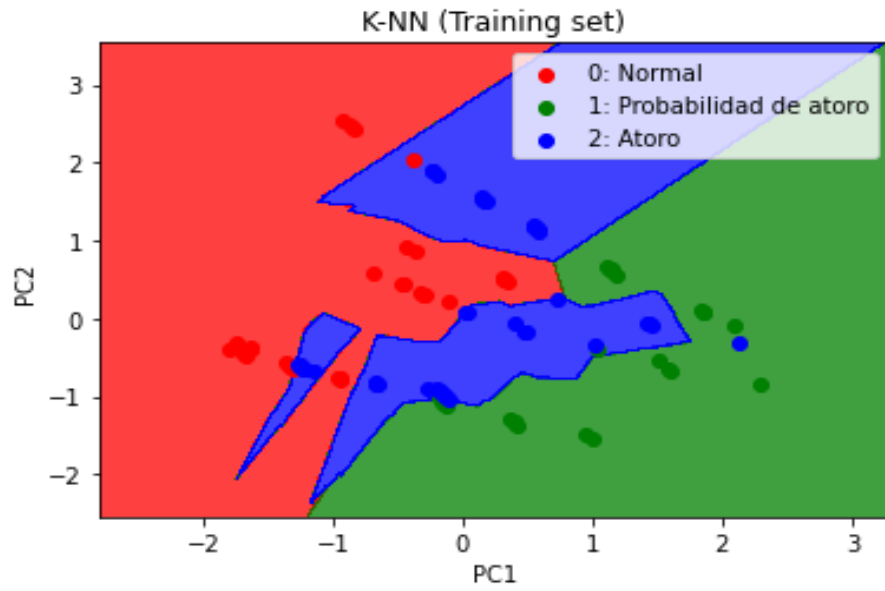


Figura 2a Gráfica del análisis de Regresión K-NN (Training set) del Escenario 2

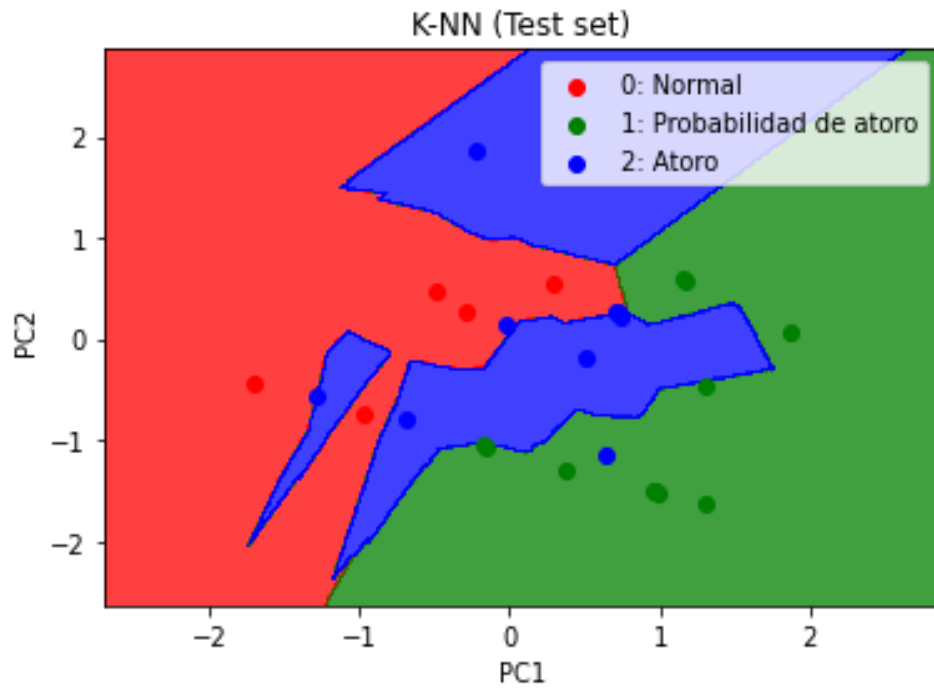


Figura 2b Gráfica del análisis de Regresión K-NN (Test set) del Escenario 2

c) Regresión SVM:

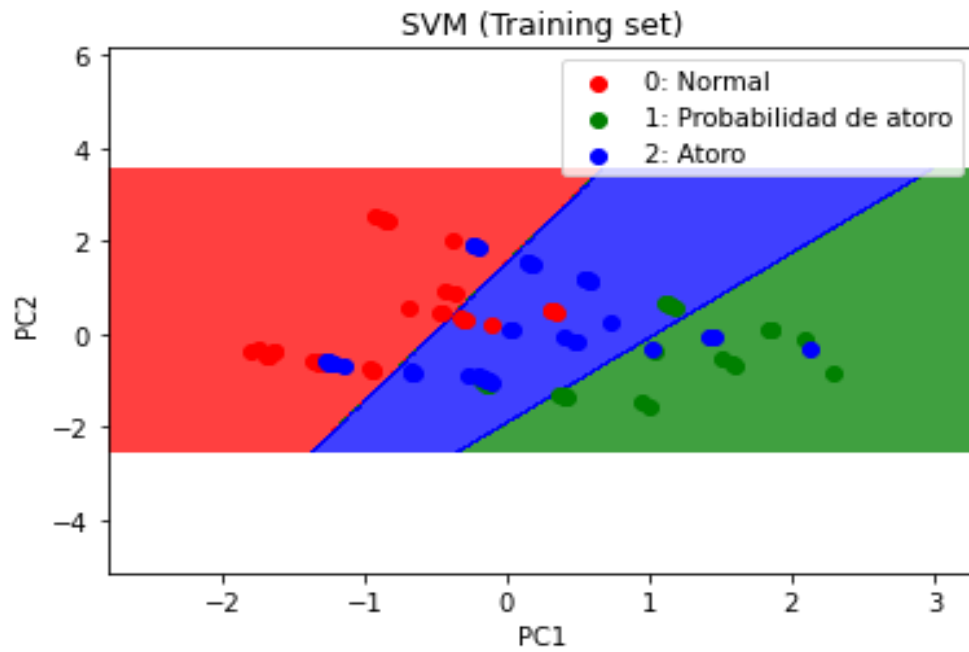


Figura 3a Gráfica del análisis de Regresión SVM (Training set) del Escenario 2

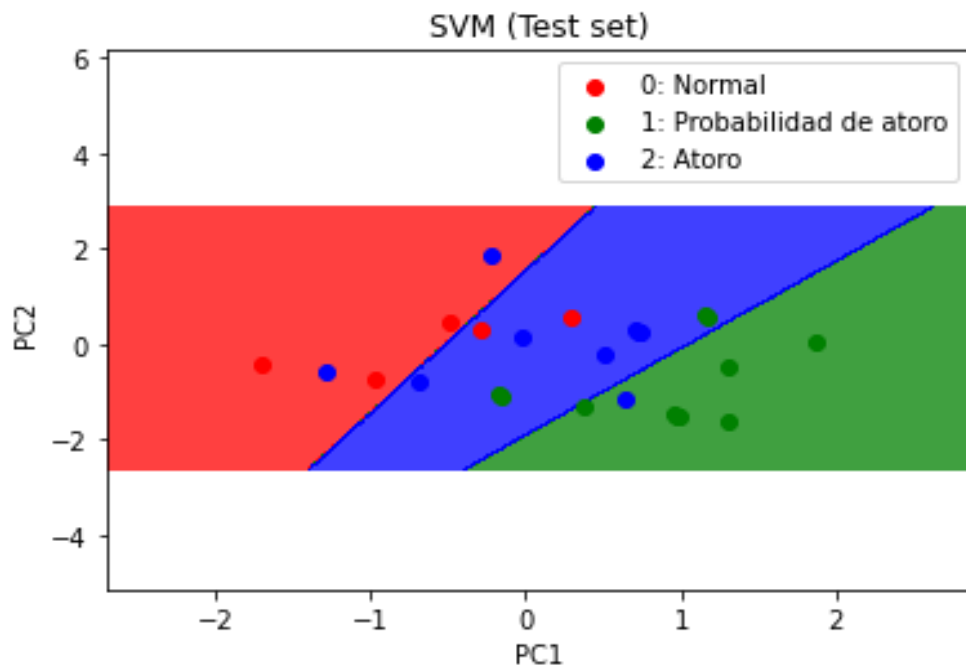


Figura 3b Gráfica del análisis de Regresión SVM (Test set) del Escenario 2.

d) Regresión Decision Tree:

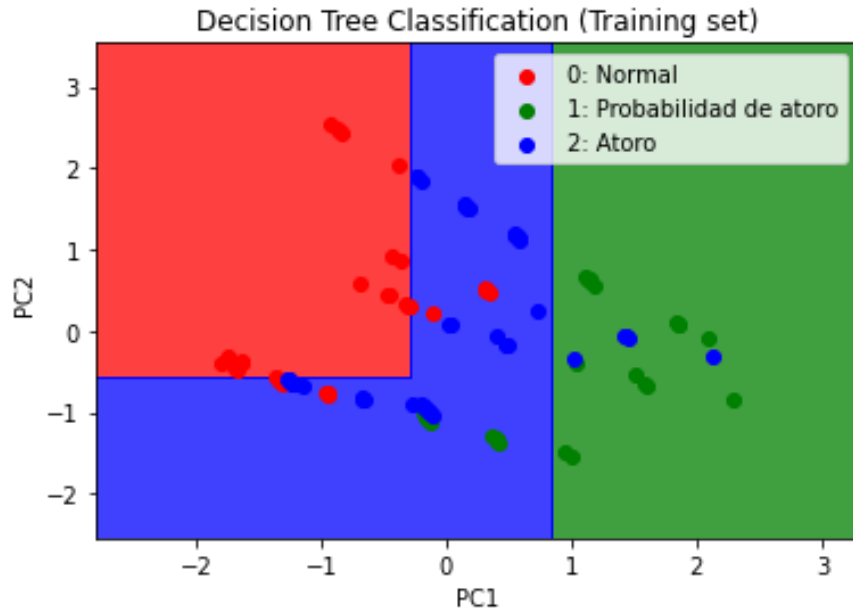


Figura 4a Gráfica del análisis de Regresión Decision Tree (Training set) del Escenario 2.

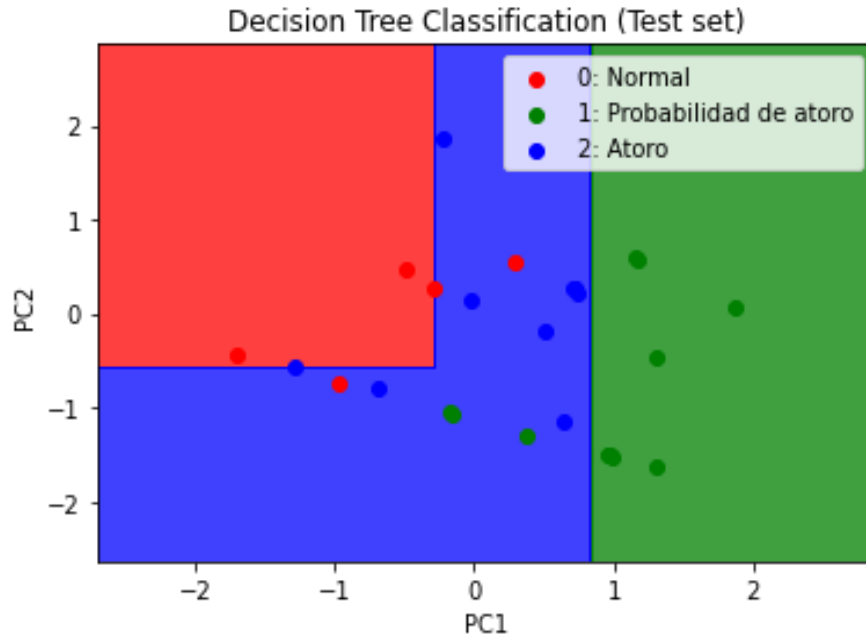


Figura 4b Gráfica del análisis de Regresión Decision Tree (Test set) del Escenario 2.