



**UNIVERSIDAD MAYOR DE SAN SIMÓN**  
**FACULTAD DE CIENCIAS Y TECNOLOGÍA**

**UNIVERSIDAD  
MAYOR DE SAN SIMÓN**  
Ciencia y Conocimiento desde 1832

**DIRECCIÓN DE POSGRADO**

**DCG**  
DIRECCIÓN DE POSGRADO  
FACULTAD DE CIENCIAS Y TECNOLOGÍA

**DIPLOMADO ESTADÍSTICA APLICADA A LA  
TOMA DE DECISIONES  
SEGUNDA VERSIÓN**

**MODELO DE CLASIFICACIÓN DEL IMPACTO DEL  
PROGRAMA INCLUSIÓN DIGITAL Y  
DISPONIBILIDAD DE RECURSOS E  
INFRAESTRUCTURA EN LAS ESCUELAS  
PÚBLICAS DE BOLIVIA UTILIZANDO MACHINE  
LEARNING**

**PROYECTO PRESENTADO PARA OBTENER EL GRADO DE LICENCIATURA EN  
INGENIERÍA ELECTRÓNICA  
MODALIDAD DOBLE TITULACIÓN**

**POSTULANTE : ARIEL ROBLES VASQUEZ  
TUTOR : ING. BRYAN OROSCO MONTAÑO**

**Cochabamba – Bolivia  
2023**



# **MODELO DE CLASIFICACIÓN DEL IMPACTO DEL PROGRAMA INCLUSIÓN DIGITAL Y DISPONIBILIDAD DE RECURSOS E INFRAESTRUCTURA EN LAS ESCUELAS PÚBLICAS DE BOLIVIA UTILIZANDO MACHINE LEARNING**

Por

Ariel Robles Vásquez

El presente documento, Trabajo de Grado es presentado a la Dirección de Posgrado de la Facultad de Ciencias y Tecnología en cumplimiento parcial de los requisitos para la obtención del grado académico de Licenciatura (o sólo diplomado) en Ingeniería Electrónica, modalidad Doble Titulación, habiendo cursado el Diplomado “Estadística Aplicada a la Toma de Decisiones” propuesta por el Centro de Estadística Aplicada (CESA) en su segunda versión.

ASESOR/TUTOR

Ing. Bryan Orosco Montaño

COMITÉ DE EVALUACIÓN

Lic. Patiño Tito Ronald Edgar, M.Sc. (Presidente)  
Ing. Guillen Salvador Roxana, M.Sc. (Coordinador)  
Ing. Espinoza Orosco José, M.Sc. (Tribunal)  
Ing. Trujillo Araníbar Wilson, M.Sc (Tribunal)



**DIRECCIÓN DE POSGRADO, FACULTAD DE CIENCIAS Y TECNOLOGÍA**  
Cochabamba, Bolivia

### **Aclaración**

Este documento describe el trabajo realizado como parte del programa de estudios de Diplomado “Estadística Aplicada a la Toma de Decisiones” en el Centro de Estadística Aplicada CESA y la Dirección de Posgrado de la Facultad de Ciencias y Tecnología. Todos los puntos de vista y opiniones expresadas en el mismo son responsabilidad exclusiva del autor y no representan necesariamente las de la institución.

---

## **Resumen**

---

La Agencia de Gobierno Electrónico y Tecnologías de Información y Comunicación (AGETIC) lanza el programa Inclusión Digital, donde por medio de capacitaciones directores, profesores y alumnos de nivel secundario, adquieren habilidades tecnológicas a través de manejo de herramientas de Software Libre, uso de TIC, herramientas de Google para que puedan aplicarlas en los métodos de enseñanza-aprendizaje y así ejercer su Ciudadanía Digital, (AGETIC, [www.agetic.gob.bo](http://www.agetic.gob.bo), 2019).

Hasta la fecha desde el momento del lanzamiento del programa, no se cuenta con la información sobre el impacto que tuvo el programa en las escuelas públicas de Bolivia, si estas capacitaciones lograron reducir la brecha digital entre tecnología y educación o si hubo aprovechamiento de las capacitaciones para así con los resultados ir mejorando el programa cada día más.

Como objetivo principal ante lo anterior, se busca crear un modelo de clasificación del impacto que tuvo el programa de inclusión digital en las escuelas públicas, contemplando la infraestructura y la enseñanza, para verificar el aprovechamiento de las capacitaciones a la parte docente y estudiantil.

Para esto se obtuvo un conjunto de datos con información relevante de profesores y sus respectivas escuelas que fueron participes del programa de inclusión digital, dentro de los cuales se encuentran datos personales, nivel de educación alcanzado, conocimientos tecnológicos, demanda de las escuelas, acceso a internet en las escuelas, etc. Con el uso de Google Colab un producto de Google, que permite escribir y ejecutar código de Python se realizó un análisis exploratorio de datos, con el fin de extraer toda información. Con la ayuda de algoritmos de clasificación de aprendizaje supervisado de machine learning como: árboles de decisión, máquinas de vectores de soporte (SVM), clasificación de Naive Bayes, K vecinos más cercanos (KNN), regresión logística y redes neuronales, se creó modelos de clasificación para verificar el impacto de la inclusión digital en las escuelas públicas. El uso de la herramienta visual como ser Tableau fue muy provechoso para visualizar los datos y ayudar a llegar a los objetivos.

Con los valores de las métricas de evaluación: Accuracy igual a 0,985; Precision igual a 0,974; Recall igual a 1,0; F1-score igual a 0,987; AUC-ROC igual a 0,999; el modelo de clasificación del algoritmo “Redes Neuronales (ANN)”, llega a ser el mejor modelo para la clasificación del impacto de la inclusión digital.

El 66,8% de los maestros mencionan tener la necesidad de contar con internet en sus unidades educativas. Y con los valores predichos como correctos y alto impacto de las escuelas en los distintos departamentos con el modelo de clasificación, el departamento de Tarija con 43 (24,86%) escuelas con calificación de alto impacto logró ponerse como el mejor con respecto a la implementación de la inclusión digital, seguidos por el departamento de La Paz con 35 (20,23%), y solamente el 8,67% para Cochabamba.

### **Palabras clave**

Impacto de inclusión digital, brecha digital, infraestructura y enseñanza, Machine learning.

*A mis padres,*

*A ustedes, que me han apoyado incondicionalmente en todo momento. A ustedes, que han estado a mi lado desde el principio, incluso cuando no creía en mí mismo. A ustedes, que me han enseñado a ser perseverante y a nunca rendirme.*

*Este trabajo es para ustedes, un pequeño homenaje a todo lo que han hecho por mí. Es el resultado de su amor, su apoyo y su fe en mí.*

## Agradecimientos

---

*Agradezco a mi esposa, Jessica Lazarte Cruz, por su amor, apoyo y comprensión. Su compañía y aliento han sido fundamentales para el éxito de este proyecto.*

*A mis padres, y toda mi familia, les agradezco por su amor, apoyo y educación. Me han inculcado el amor por el aprendizaje y la investigación, y me han dado las herramientas necesarias para alcanzar mis metas.*

*Agradezco a mi tutor, el Ing. Bryan Orozco Montaño, por su apoyo incondicional y su orientación durante la realización de este trabajo. Su experiencia y conocimientos me han sido de gran ayuda para alcanzar los objetivos de esta investigación.*

*A todo el equipo de profesionales de la carrera de ingeniería electrónica por la transmisión de sus conocimientos y por brindarme la información generada y utilizada en el presente trabajo.*

## Tabla de contenidos

---

1. Introducción .....	1
1.1. Antecedentes .....	2
1.2. Justificación .....	3
1.3. Planteamiento del problema .....	4
1.4. Objetivo general.....	4
1.4.1. Objetivos específicos .....	4
1.4.2. Alcances y limitaciones .....	5
2. Marco teórico .....	6
2.1. Inclusión digital.....	6
2.1.1. Barreras de la inclusión digital. ....	6
2.1.2. Gobierno y empresas móviles ante la inclusión digital.....	7
2.1.3. Bajo nivel de demanda del servicio de internet (antes de COVID-19). ....	7
2.1.4. COVID-19 ante la educación y la inclusión digital.....	8
2.1.5. Brecha e inclusión digital.....	8
2.1.6. Inclusión digital en Bolivia. ....	9
2.2. Machine Learning. ....	10
2.2.1. Inteligencia Artificial.....	10
2.2.2. Exploración de Machine Learning .....	10
2.2.3. Aprendizaje Supervisado. ....	11
2.2.4. Aprendizaje No Supervisado. ....	11
2.3. Modelos de Machine Learning.....	11
2.3.1. Modelos Lineales.....	11
2.3.2. Modelos de árbol.....	12
2.3.3. Redes Neuronales.....	12
2.4. Fases de desarrollo del modelo de machine learning.....	12
2.4.1. Fase de entrenamiento.....	12
2.4.2. Fase de prueba.....	12
2.4.3. Fase de evaluación.....	12
2.5. Ámbitos de la aplicación de algoritmos machine learning. ....	14
2.6. Metodología CRISP-DM. ....	14

---

2.6.1	Comprendión del negocio.....	15
2.6.2	Comprendión de los datos.....	15
2.6.3	Preparación de los datos.....	16
2.6.4	Modelado. ....	17
2.6.5	Evaluación.....	17
2.6.6	Despliegue o implantación.....	17
2.7.	Google Colaboratory.....	18
2.7.1	Características y funcionamiento de Google Colab.....	18
2.7.2	Utilidad de Google Colab.....	19
2.8.	Tableau.....	19
2.8.1	Función de Tableau.....	19
2.8.2	Tableau como analizador visual.....	20
2.9.	Power BI.....	20
2.10.	Clasificación de impacto de inclusión digital.....	21
3.	Metodología.....	25
3.1.	Área de estudio .....	25
3.2.	Flujograma metodológico .....	26
3.2.1.	Etapas de desarrollo.....	27
3.2.2.	Clasificación de impacto.....	28
3.3.	Fuentes de información .....	28
3.3.1.	Fuente de información secundaria .....	28
3.4.	Aplicación del Metodología CRISP-DM.....	29
3.4.1.	Comprendión de Negocio. ....	29
3.4.2.	Comprendión de Datos.....	29
3.4.3.	Preparación de los Datos.....	34
3.4.4.	Modelado. ....	59
4.	Resultados y Discusión.....	74
4.1	Resultados de la exploración y análisis del conjunto de datos. ....	74
4.2	Resultados del modelo de clasificación de aprendizaje supervisado de machine learning.....	85
4.2.1.	Exactitud (Accuracy).....	85
4.2.2.	Precisión Positiva (Precision).....	86
4.2.3.	Recall (Sensibilidad o Exhastividad).....	86
4.2.4.	F1-Score. ....	86

4.2.5. AUC-ROC (Área bajo la curva ROC).....	86
4.2.6. Curva ROC (Receiver Operating Characteristic).....	86
4.3 Elección del modelo para clasificación del impacto del programa de inclusión digital.....	87
4.4 Discusión de resultados de análisis del conjunto de datos.....	90
5. Conclusiones.....	95
6. Recomendaciones.....	97
Referencias bibliográficas.....	98
Anexos.....	100
Anexo 1. Tabla resumen de la encuesta sobre la inclusión digital a maestros de las escuelas públicas de Bolivia.....	100
Anexo 2. Gráficas de variables cuantitativas.....	101
Anexo 3. Gráficas de variables cualitativas.....	102
Anexo 3.1: Grafico de Barras sobre la frecuencia de los profesores por Departamento.....	102
Anexo 3.2: Grafico de Barras sobre el nivel académico de los profesores.....	102
Anexo 3.3: Gráfica de frecuencia de uso de redes sociales por los profesores.....	103
Anexo 3.4: Grafica de frecuencia de la utilidad de la computadora por los profesores.....	103
Anexo 3.5: Grafica de frecuencia de la utilidad del uso de celular por los profesores.....	103
Anexo 3.6: Grafica de frecuencia de la utilidad del uso de Internet por los profesores.....	104
Anexo 3.7: Grafica de frecuencia del uso de conocimientos informáticos, ofimáticos y herramientas de Google.....	104
Anexo 3.8: Grafica de capacitación, conocimiento e implementación de métodos TIC.....	105
Anexo 3.9: Grafica de la frecuencia de implementación de TIC en la enseñanza de los profesores.....	105
Anexo 3.10: Grafica de la frecuencia del modo de preparación de los estudiantes para sus clases.....	106
Anexo 3.11: Grafica de frecuencia de elaboración del material de enseñanza de los profesores ..	106
Anexo 3.12: Grafica de la frecuencia de la modalidad de Clases en las enseñanzas de los profesores.....	107
Anexo 3.13: Grafica de la frecuencia de los requerimientos de infraestructura de las escuelas públicas de Bolivia.....	107
Anexo 3.14: Grafica de la calificación que tuvo el proyecto de inclusión digital.....	108
Anexo 3.15: Grafica de la frecuencia que determina si el uso de tecnología mejoró el aprendizaje en los estudiantes.....	108
Anexo 4: Algoritmo Logistic Regression como Modelo de clasificación.....	109
Anexo 5: Algoritmo Gaussian como modelo de clasificación.....	111

---

Anexo 6: Algoritmo KNeighbors como modelo de clasificación.....	113
Anexo 7: Modelo de clasificación del algoritmo SVM o SVC.....	114
Anexo 8: Modelo de clasificación del algoritmo Redes Neuronales ANN.....	116
Anexo 9: Algoritmo Decision Tree como modelo de clasificación.....	118
Anexo 10: Algoritmo Redes Neuronales DNN como modelo de clasificación.....	120
Anexo 11: Encuesta a los profesores, para ayuda a definir variables de clasificación.....	122
Anexo Principal CD.....	123

## **Lista de figuras**

---

Figura 2-1: Matriz de confusión.....	13
Figura 2-2: Fases de la metodología CRISP-DM.....	15
Figura 2-3: Interfaz de Google Colab.....	18
Figura 2-4: Interfaz de Tableau.....	20
Figura 2-5: Conexión y funcionamiento de Power Bi.....	21
Figura 3-1: Mapa de Bolivia.....	25
Figura 3-2: Flujograma metodológico.....	26
Figura 3-3: Conjunto de datos “Profesores_fin” en formato csv.....	30
Figura 3-4: Configuración Inicial en Google Colab.....	35
Figura 3-5: Comando head() para visualizar el DataFrame.....	36
Figura 3-6: Comando info() para ver las cantidades .....	36
Figura 3-7: Comando shape.....	36
Figura 3-8: Comando dtypes.....	36
Figura 3-9: Estadística descriptiva de los datos.....	37
Figura 3-10: Código Python para el Histograma de Edad.....	37
Figura 3-11: Histograma de Edad.....	38
Figura 3-12: Histograma de Años de experiencia.....	38
Figura 3-13: Diagrama de barras de la variable Departamento.....	39
Figura 3-14: Diagrama de barras de Nivel Académico .....	39
Figura 3-15: Diagrama de barras de Materias dictadas por los maestros.....	40
Figura 3-16: Diagrama de barras de Redes Sociales.....	40
Figura 3-17: Diagrama de barras de Acceso a internet .....	41
Figura 3-18: Diagrama de barras de usos de conocimientos Informáticos .....	42
Figura 3-19: Diagrama de barras de Incorporación de TIC.....	43
Figura 3-20: Diagrama de Modalidades de Clases.....	43
Figura 3-21: Diagrama de barras de Requerimientos de los Colegios.....	44
Figura 3-22: Diagrama de barras de Requerimientos de los Colegios.....	45
Figura 3-23: Verificación de valores nulos.....	45
Figura 3-24: Mapa de Calor .....	46
Figura 3-25: Valores Nulos.....	46

---

Figura 3-26: Tratamiento de Valores Nulos.	47
Figura 3-27: Eliminación de Valores Atípicos.	47
Figura 3-28: Diagrama de cajas de Edad.	48
Figura 3-29: Diagrama de cajas de Años de experiencia.	48
Figura 3-30: Diagrama de días en la semana usando internet.	49
Figura 3-31: Diagramas de correlación Edad Vs Años de Experiencia.	49
Figura 3-32: Código de matriz de correlación.	50
Figura 3-33: Matriz de Correlación.	50
Figura 3-34: Análisis bivariado de variables categóricas.	51
Figura 3-35: Análisis bivariado de variables cualitativas.	51
Figura 3-36: Análisis bivariado de variables cualitativas.	52
Figura 3-37: Análisis bivariado de variables cualitativas.	52
Figura 3-38: Análisis bivariado de requerimientos de los colegios y la existencia de un aula tecnológica.	53
	53
Figura 3-39: Análisis bivariado de variables cualitativas.	54
Figura 3-40: Análisis bivariado de variables cualitativas.	55
Figura 3-41: Análisis bivariado de variables cualitativas.	55
Figura 3-42: Análisis bivariado de variables cualitativas.	56
Figura 3-43: Creación de Nuevo DataFrame con variables sobresalientes.	57
Figura 3-44: Creación de la variable objetivo.	57
Figura 3-45: Vista del nuevo DataFrame.	57
Figura 3-46: Desequilibrio en variable objetivo.	58
Figura 3-47: Equilibrio de la variable objetivo.	58
Figura 3-48: Conjunto de Datos listo para el Modelo.	59
Figura 3-49: Preparación, entrenamiento y prueba del Modelo en Regresión Logística Binaria.	60
Figura 3-50: Métricas de evaluación y matriz de confusión del modelo en Regresión Logística Binaria.	60
Figura 3-51: Valores de métricas de evaluación del modelo de Regresión Logística Binaria.	61
Figura 3-52: Modelo en Maquinas de Vectores de Soporte (SVM).	62
Figura 3-53: Resultado del modelo en Maquinas de Vectores de Soporte (SVM).	62
Figura 3-54: Valores de métricas de evaluación del modelo Maquinas de Vectores de Soporte (SVM).	63
Figura 3-55: Modelo en K Vecinos más cercanos (K-NN).	64
Figura 3-56: Resultados del modelo en K Vecinos más cercanos (K-NN).	65

Figura 3-57: Modelo en Árbol de Decisiones.....	66
Figura 3-58: Resultados del modelo en Árbol de Decisiones .....	66
Figura 3-59: Valores de métricas de evaluación del modelo árbol de decisiones.....	67
Figura 3-60: Modelo en Naive Bayes.....	68
Figura 3-61: Resultado del modelo en Naive Bayes.....	68
Figura 3-62: Valores de métricas de evaluación del modelo Naive Bayes.....	69
Figura 3-63: Modelo en Redes Neuronales profundas (ANN).....	70
Figura 3-64: Resultado del modelo en Redes Neuronales profundas (ANN).....	71
Figura 3-65: Valores de métricas de evaluación del modelo Redes Neuronales (ANN).....	71
Figura 3-66: Exportación del modelo entrenado. ....	72
Figura 3-67: Importación del modelo ya entrenado.....	72
Figura 3-68: DataFrame del nuevo conjunto de datos. ....	72
Figura 3-69: Uso del modelo con el nuevo conjunto de datos.....	73
Figura 3-70: Unión de predicciones y conjunto de datos nuevos.....	73
Figura 4-1: Nivel académico de los maestros en las escuelas públicas.....	74
Figura 4-2: Uso de redes sociales en los maestros de las escuelas públicas.....	75
Figura 4-3: Resultado de mayoría de edad en los maestros.....	76
Figura 4-4: Uso del celular.....	76
Figura 4-5: Uso de la computadora.....	77
Figura 4-6: Uso del internet en los profesores.....	78
Figura 4-7: Preparación del material de los maestros. ....	79
Figura 4-8: Resultado de preparación de los estudiantes.....	80
Figura 4-9: Resultado de conocimientos en herramientas de Google.....	80
Figura 4-10: Resultado de conocimientos básicos en informática.....	81
Figura 4-11: Resultado de implementación de TIC en los colegios. ....	82
Figura 4-12: Resultado de requerimiento de los colegios.....	82
Figura 4-13: Impacto de inclusión digital de tipo Alto en los 9 departamentos de Bolivia. ....	83
Figura 4-14: Impacto de inclusión digital de tipo Alto en los 9 departamentos de Bolivia. ....	84
Figura 4-15: Impacto alto y bajo en los 9 departamentos de Bolivia. ....	84
Figura 4-16: Porcentaje de inclusión digital según la cantidad de participación de profesores. ....	85
Figura 4-17: Matriz de confusión de algoritmos de clasificación en aprendizaje supervisado. ....	87
Figura 4-18: Curva ROC (Redes Neuronales ANN). ....	88

---

Figura 4-19: Valores predichos del modelo (Redes Neuronales ANN).....	89
Figura 4-20: Resultados de predicción del nuevo conjunto de datos que se usó como prueba del modelo.....	89
Figura 4-21: Espacios utilizados por los estudiantes para acceder a internet.....	90
Figura 4-22: Lugares de acceso a internet con más frecuencia.....	90
Figura 4-23: Estudiantes y usos de Internet.....	91
Figura 4-24: Utilidad del internet .....	91
Figura 4-25: Recursos elaborados por los profesores .....	92
Figura 4-26: Elaboración del material de clases.....	92
Figura 4-27: Matriz de Confusión Predicción del Modelo en Parques. ....	93
Figura 4-28: Descripción de Resultados Predicción del Modelo en Parques.....	93
Figura 4-29: Matriz de evaluación del modelo de clasificación Decision Tree. ....	94
Figura 4-30: Resultado de Métricas del Modelo de Predicción en Parques.....	94

## **Lista de tablas**

Tabla 2-1: Características que catalogan una buena escuela con inclusión digital .....	23
Tabla 2-2: Tabla con variables fundamentales de las encuestas.....	23
Tabla 2-3: Valores calificativos para determinar alto y bajo impacto.....	24
Tabla 3-1: Encuesta Final-Profesores de Inclusión Digital.....	28
Tabla 3-2: Estrategia de Negocios .....	29
Tabla 3-3: Conjunto de datos “Profesores_final” en formato xlsx.....	30
Tabla 3-4: Tabla de datos personales de los profesores .....	31
Tabla 3-5: Tabla de datos de nivel académico de los profesores.....	31
Tabla 3-6: Tabla de datos de materias que imparten los profesores .....	31
Tabla 3-7: Tabla de herramientas de apoyo de educación de los profesores .....	32
Tabla 3-8: Tabla que muestra la utilidad de las herramientas como computadora y celular.....	32
Tabla 3-9: Tabla de acceso y utilidad del internet en los profesores .....	33
Tabla 3-10: Tabla de uso de redes sociales de los profesores .....	33
Tabla 3-11: Conocimientos e implementaciones de los profesores para la educación.....	33
Tabla 3-12: Tabla de formas de preparación de los maestros y estudiantes.....	34
Tabla 3-13: Tabla de requerimientos de las escuelas.....	34
Tabla 3-14: Tabla de características de los colegios .....	34
Tabla 3-15: Comandos de Python para exploración inicial de Datos.....	35
Tabla 3-16: Descripción del significado de las variables saber 1,2,3,4,5,6,7 .....	42
Tabla 4-1: Descripción de las variables saber 1,2,3,4,5,6,7.....	81
Tabla 4-2: Métricas de algoritmos.....	86
Tabla 4-3: Métricas del algoritmo Redes Neuronales (ANN).....	88
Tabla 4-4: Matriz de confusión del algoritmo Redes Neuronales (ANN).....	88
Tabla 4-5: Tabla de las métricas del modelo de clasificación con algoritmo Redes Neuronales (ANN).94	

## 1. Introducción.

En los últimos años se ha podido observar la creciente evolución de la tecnología y la inteligencia artificial en todo el mundo. El uso de herramientas tecnológicas y tecnologías de información y comunicación (TIC) de última generación en toda la sociedad se ha universalizado, modificando sustancialmente las relaciones económicas, sociales, políticas y culturales. También se han generado nuevos paradigmas en la relación Estado y Sociedad a partir de la implementación y uso de la tecnología (Agetic, 2018).

Ante esto Bolivia con la ayuda de AGETIC (Agencia de Gobierno Electrónico y Tecnologías de Información y Comunicación) estuvo lanzando el programa llamado inclusión digital dirigido a las escuelas públicas con el fin de poder capacitar a los estudiantes, profesores y directores en relación a conocimientos sobre herramientas tecnológicas y TIC (Tecnologías de la Información y las Comunicaciones), e implementar tecnologías digitales en sus aulas. Después de la última capacitación que realizó el gobierno de Bolivia en el año 2021, se realizó una encuesta al azar a varios profesores de escuelas públicas de todos los departamentos con el fin de recabar información sobre; datos personales de los profesores, de los colegios, su infraestructura, métodos de enseñanza y otras características (AGETIC, [www.digital.gob.bo](http://www.digital.gob.bo), 2019).

Con el uso del conjunto de datos obtenida de la encuesta a varios profesores de escuelas públicas de los diferentes departamentos de Bolivia, que fueron obtenidos de la página oficial [datos.gob.bo](http://datos.gob.bo) con el fin de realizar un Análisis Exploratorio de Datos, para extraer información actualizada sobre; el impacto que tuvo el programa en la educación y los métodos de enseñanza que usan los maestros hoy en día. Mismo que ayudara a verificar si las escuelas públicas de Bolivia cuentan con ambientes tecnológicos adecuados y equipamiento correspondiente en donde los estudiantes y maestros puedan hacer el uso e implementación de las herramientas tecnológicas y las TIC; a su vez si la inclusión digital pudo mejorar la educación y la forma clásica de aprender y enseñar.

Para poder clasificar a escuelas con alto y bajo impacto sobre la inclusión digital, se usará algoritmos de clasificación de aprendizaje supervisado de machine learning, como árboles de decisión, máquinas de vectores de soporte (SVM), clasificación de Naive Bayes, K vecinos más cercanos (KNN), regresión logística y redes neuronales. También el uso de Tableau y Power Bi como herramienta de visualización de datos.

Sacar unas buenas conclusiones sobre el estado actual de la educación después del último lanzamiento del programa de inclusión digital, ayudará a tomar decisiones futuras para así mejorar el programa de AGETIC.

### **1.1. Antecedentes.**

La enseñanza en las escuelas públicas en estos últimos años ha sido un enfoque primordial y de estudio en el mundo entero. En Bolivia particularmente se ha visto reforzada mucho por el programa de inclusión digital, el cual busca a través de la alfabetización tecnológica, desarrollar habilidades digitales en grupos clave de la sociedad boliviana, introduciendo proyectos que permitan romper las brechas digitales, además de abrir accesos y oportunidades para un mejor desarrollo en el cotidiano vivir.

Realizando capacitaciones a los estudiantes y enseñando a manejar; programas, sitios web, herramientas de aprendizaje virtual, plataformas educativas y otros, a su vez la misma acción para con los profesores y directores para que puedan usar estas herramientas e implementar en sus métodos de enseñanza; se ha tratado de que tanto como estudiantes y profesores puedan ir siempre de la mano con el avance tecnológico y así poder ir mejorando cada día más la educación. (AGETIC, 2019) (BRECHA CERO, 2019).

En la actualidad, los avances tecnológicos y su alta velocidad de cambio han supuesto un reto para las organizaciones que deben responder ante ellos y anticiparse a nuevas transformaciones. Uno de esos avances es la Inteligencia Artificial (IA), que está generando desarrollos sorprendentes e impactando en diferentes sectores económicos como, por ejemplo, industria, servicios, gestión pública, medicina, medio ambiente, agricultura, energía, transporte y movilidad, educación, finanzas, entre otros., con procesos innovadores a partir del uso y procesamiento masivo de datos, y que han repercutido en la formación y cualificación en habilidades digitales de todos los actores involucrados. Esta situación conlleva a la comprensión y exploración de las subáreas actuales de la IA y cómo, a partir de la generación de información, se convierten en facilitadoras para la toma de decisiones.

Una de esas subáreas es el Machine Learning (ML) o aprendizaje automático, que puede ayudar a las compañías a transformar y comprender sus datos a partir de la elaboración de modelos para la detección de patrones y su aplicación en el entorno real, utilizando las herramientas tecnológicas adecuadas para ello, aplicando y comprendiendo sus diferentes algoritmos e interpretando apropiadamente los resultados mediante sus métricas (Francisco Javier Lena Acebo, 2021).

Una de las características más salientes de la tercera revolución industrial, y que ha dado origen a la sociedad de la información (Rifkin, 2011) ha sido el desarrollo de la tecnología digital. Los mayores hitos están relacionados con la aparición y desarrollo de elementos como los ordenadores personales o internet, que han transformado la forma de comunicación humana y la educación (Elosua, 2021).

La forma en que nos comunicamos está en constante evolución. Las tecnologías de la información y de la comunicación (TIC) y las redes sociales intervienen directamente en la cultura de un país. En México, el número de usuarios ha pasado ya los 79 millones. Sin embargo, la desigualdad en el acceso a las TIC, impide el propósito de agilizar la relación sociedad-red social que permite un aprendizaje interactivo y dinámico en la educación (Guadalupe Aurora Maldonado Berea, 2019).

Internet se constituye en una de las principales fuentes de herramientas y conocimientos que contribuyen al desarrollo socioeconómico y cultural de las naciones en los últimos años. En tal sentido, una idónea infraestructura de Tecnologías de Información y Comunicación (TIC) y telecomunicaciones resulta decisiva para el desarrollo de políticas públicas de educación, salud, seguridad ciudadana y desarrollo económico de alta calidad en el país. Por estos motivos, la infraestructura debe apuntar a garantizar un uso generalizado, intensivo e íntegro de Internet; sin embargo, la infraestructura de Telecomunicaciones y TIC en Bolivia presenta problemas como el elevado costo, la baja calidad y el uso limitado o restringido del servicio de Internet (Bustillos, 2016).

Las TIC han pasado de ser un privilegio de algunos a ser una necesidad para la población y el desarrollo del Estado, convirtiéndolas en servicios básicos que deberían ser de acceso universal. Según este precepto, que la labor de los Estados modernos es el de utilizar y otorgar los medios necesarios para el desarrollo de las TIC, ya sea mediante inversiones públicas o privadas que busquen como fin último la innovación en el aparato productivo y un mayor acceso a las TIC, que trae consigo mejoras en la educación (LLANOS, 2019).

## **1.2. Justificación.**

La realización de este proyecto consiste en un Análisis Exploratorio de Datos (AED) del conjunto de información de los profesores y sus escuelas públicas, y también la creación de un modelo que ayuda a clasificar el impacto alto y bajo de las unidades educativas con respecto a la inclusión digital; es de mucho beneficio a la educación, ya que ayuda a identificar si las escuelas públicas cuentan con acceso a internet, aulas tecnológicas, computadoras para los maestros como apoyo para la educación, el uso de TIC en sus métodos de enseñanza, que nivel de conocimientos de ofimática e informática tienen los profesores y áreas en las que el programa inclusión digital lanzado por AGETIC necesita ir mejorando en cuanto a capacitación y orientación tanto a estudiantes y profesores; también permite ver el estado actual de la infraestructura tecnológica de las escuelas a nivel nacional, y si estas cuentan con herramientas y recursos necesarios para que los profesores imparten clases tecnológicas con total normalidad y poder añadir estos métodos de enseñanza a sus clases.

Ayuda a determinar en qué departamentos se encuentran las escuelas con impacto bajo en inclusión digital, también los que tienen menos apoyo y recursos para la implementación de TIC en la enseñanza y así se da a conocer esta situación a AGETIC, para poder ayudar a tomar decisiones a las autoridades correspondientes, como solicitar los recursos que son necesarios para alcanzar el nivel adecuado en la enseñanza, ya que se tiene datos sobre las necesidades de mejoramiento y las herramientas que necesitan contar para implementar conocimientos tecnológicos en aula, con la cual las escuelas públicas tendrán una mejora tecnológica y comunicación con el mundo. Se verifica el impacto que tuvo la inclusión digital en las escuelas públicas y la educación, lo cual puede conducir a una mejor asignación de recursos para el programa y una mayor eficiencia del mismo.

### **1.3. Planteamiento del problema.**

En el año 2021, se realizó el último lanzamiento del programa de inclusión digital para las escuelas públicas con el fin de eliminar la brecha digital entre la tecnología y las instituciones de educación. Desde entonces no se cuenta con una información actualizada sobre el impacto, reducción de la brecha digital o del aprovechamiento al máximo del programa, no se cuenta con información que nos ayude a verificar si hubo un impacto alto o bajo en la educación de las escuelas públicas de Bolivia.

No se cuenta con información de los maestros si utilizan al máximo las capacitaciones tecnológicas en el manejo y uso de TIC, herramientas de Google (Drive, Académico), conocimientos de ofimática e informática, también si se sienten listos para incluir o si ya incluyeron estas capacitaciones y herramientas a sus métodos de enseñanza, o si hubo mejora en el aprendizaje de los estudiantes después de la capacitación, o si las escuelas cuentan con aulas tecnológicas para realizar estas implementaciones, si se tienen herramientas como computadoras, pizarrones tecnológicos, uso de almacenamiento de documentos en la nube, acceso a clases virtuales, acceso a clases grabadas anteriores de las materias, o si las unidades educativas cuentan con acceso a internet, ya que el no tener acceso y no usar los puntos anteriores descritos, no se estaría aprovechando en sus totalidad toda la capacitación que se realizó con el programa de inclusión digital.

Dado la descripción anteriormente mencionada, se ve la necesidad de realizar un Análisis Exploratorio de Datos (AED) y un modelo que ayude a clasificar un impacto alto y bajo de la inclusión digital en la educación, para identificar la situación actual de la educación con respecto al uso de tecnologías en aulas y poder ayudar a tomar decisiones de mejoramiento a aquellos departamentos que tengan impacto bajo, dando a conocer la situación a las autoridades correspondientes encargados de la educación, y también a AGETIC para así mejorar el programa de inclusión digital que sea lanzado en un futuro, todo esto para contribuir al mejoramiento de la educación en las escuelas públicas de Bolivia.

### **1.4. Objetivo general**

Realizar un modelo de clasificación de impacto del programa inclusión digital según la infraestructura y enseñanza en las escuelas públicas utilizando machine learning, para apoyo de toma de decisiones del programa AGETIC y la educación boliviana.

#### **1.4.1. Objetivos específicos**

- Recopilar la información primaria y secundaria de maestros y sus escuelas públicas.
- Realizar un Análisis Exploratorio de Datos, del conjunto de información obtenida, para descubrir relación entre variables, características sobresalientes e información valiosa.

- Realizar el modelado, entrenamiento y prueba del modelo de clasificación de aprendizaje supervisado de machine learning, para determinar el tipo de impacto que tuvo la inclusión digital.
- Evaluar los diferentes modelos de clasificación como árboles de decisión, máquinas de vectores de soporte (SVM), clasificación de Naive Bayes, K vecinos más cercanos (KNN), regresión logística y redes neuronales, para determinar cuál es el más apropiado.

#### **1.4.2. Alcances y limitaciones**

Los alcances y limitaciones del estudio propuesto son los siguientes:

- Proporcionará información sobre el uso de la tecnología en las escuelas públicas de Bolivia.
- Ofrecerá una visión más completa del impacto de la inclusión digital en las escuelas públicas de los departamentos de Bolivia.
- Identificara los factores que contribuyen y los que son perjudiciales en la educación en las escuelas públicas.
- Desarrollará un modelo de clasificación de machine learning del impacto de inclusión digital que pueda ser utilizado por otras investigaciones.
- El proyecto pretende ayudar a mejorar la educación en Bolivia.
- Los datos disponibles pueden no ser suficientes para capturar el impacto completo de la inclusión digital.
- El conjunto de datos no se encuentra equilibrado en cantidad de maestros por departamento.
- El estudio no incluye información sobre el impacto del programa de inclusión digital en el rendimiento de los estudiantes.

En general, el estudio propuesto es un proyecto que tiene el potencial de apoyar al mejoramiento de la educación en Bolivia.

## 2. Marco teórico

Las tecnologías de la información y las comunicaciones (TIC) son hoy en día un bien necesario en la vida cotidiana. Sus ventajas las convierten en recursos estratégicos para el desarrollo de las nuevas economías basadas en el conocimiento. El acceso, la disponibilidad y el uso de las TIC son temas fundamentales para el desarrollo social y económico de los pueblos. Esta nueva realidad ha demandado a los gobiernos la implementación de políticas y programas que garanticen su integración en la Sociedad de la Información y el Conocimiento. Para superar las desigualdades generadas por las TIC, se están llevando a cabo en distintos países planes de inclusión digital que van más allá del simple equipamiento tecnológico (Sulma Farfán Sossa, 2015).

### 2.1. Inclusión digital

La inclusión digital está definida como la expansión de la conectividad global, y la adopción de internet local y móvil donde proporcionan amplios beneficios sociales y económicos al ofrecer servicios de comunicaciones a la parte de la población que no está conectada, además de los posibles efectos de reducción de la pobreza, mejora en la infraestructura, servicios y aumento mayor en el acceso y uso del internet. Si se mantienen las barreras a la inclusión digital, aquellas comunidades que no están conectadas o están desatendidas corren el riesgo de quedar aún más rezagadas, haciendo que la brecha digital sea aún más profunda (Arese, 2016).

El esfuerzo de los gobiernos en eliminar las barreras creadas por la falta de alfabetización digital y de contenido local es central, además de ser instrumental en el desarrollo de aptitudes digitales, a través del financiamiento y fomento del uso de las TIC y de su aprendizaje en las escuelas. Mediante la promoción de nuevos emprendimientos e innovaciones de tecnología local, los gobiernos también pueden fomentar la creación de contenido local comprensible y pertinente a las necesidades específicas de los usuarios de internet. (Arese, 2016).

#### 2.1.1. Barreras de la inclusión digital.

Aunque las tasas de alfabetización básicas de los países latinos son más altas que el promedio mundial, sigue habiendo una brecha entre la alfabetización y las aptitudes digitales que son consideradas barreras de la inclusión digital. La falta de suficiente infraestructura y apoyo a la enseñanza de conocimientos digitales impide que muchas personas exploren los beneficios que ofrece internet. Este problema se agrava aún más cuando se considera la oferta limitada de contenido interesante, no sólo en relación al idioma local, sino también a su relevancia (Arese, 2016).

El análisis de datos de tráfico web demuestra que menos del 30% de contenido de los países latinos está en idioma local. Más aún, el contenido que se encuentra en las tiendas de aplicaciones y las páginas web

de los operadores de telefonía móvil está relacionado con entretenimiento en su mayoría. De esta manera, se crea una falsa impresión, que el internet es sólo una herramienta para entretenimiento, una percepción que encubre la verdadera importancia y el potencial de cambio de estilo de vida que ofrece internet (Arese, 2016).

### **2.1.2. Gobierno y empresas móviles ante la inclusión digital.**

Para que se logren todos los beneficios sociales y económicos impulsados por las inversiones en infraestructura de banda ancha móvil, tanto los gobiernos como las organizaciones de desarrollo y los operadores móviles deben unir sus fuerzas para hacer que el internet sea accesible, fácil de usar, útil y comprensible para todos.

Dado que las redes móviles son el principal medio de acceso a internet en la región y que la cobertura de 5G alcanzará niveles similares a la de 4G, creemos que el servicio móvil será la clave para mejorar tanto el nivel de adopción de internet como su acceso, siempre y cuando se cuente con un entorno regulatorio adecuado. Mediante la promoción de nuevos emprendimientos e innovaciones de tecnología local, los gobiernos también pueden fomentar la creación de contenido local comprensible y pertinente a las necesidades específicas de los usuarios de internet móvil, los operadores móviles pueden aprovechar sus canales de distribución y asociarse tanto con gobiernos como con otras organizaciones de desarrollo para aumentar la alfabetización digital y el conocimiento de las personas sin conexión a internet. (Arese, 2016)

### **2.1.3. Bajo nivel de demanda del servicio de internet (antes de COVID-19).**

El mercado de telefonía móvil en América Latina está madurando rápidamente, con un incremento del 5% anual promedio en suscriptores únicas durante los últimos cinco años y un crecimiento esperado de 4% anual promedio adicional esperado para los próximos cinco años. Sin embargo, se podría reducir la brecha digital aún más si los servicios de banda ancha móvil alcanzarán a una mayor parte de la población (Arese, 2016).

Más de 300 millones de personas en la región aún no cuentan con una suscripción a servicios de internet local y móvil y aún menor en el caso de la banda ancha móvil (3G y 4G), con casi 7 de cada 10 personas sin una suscripción de banda ancha móvil. Curiosamente la brecha de cobertura en América Latina y el Caribe es relativamente pequeña, con solo alrededor de 10% de la población, o 64 millones de personas, sin cobertura de redes 3G o 4G. Sin embargo, 57%, o 360 millones, de latinoamericanos que cuentan con cobertura, no utilizan la banda ancha móvil, lo cual indica una importante brecha en la demanda a pesar de la disponibilidad de redes (Arese, 2016).

Para entender los motivos de esta débil demanda, se investigaron los perfiles de las personas sin conexión a internet, dónde viven y a qué grupo socioeconómico pertenecen. Los resultados de la Encuesta al Consumidor 2015 realizada por GSMA Intelligence, combinados con los de otras encuestas nacionales

de hogares de América Latina, apuntan a tres principales barreras para la adopción de internet móvil: ausencia de contenido local relevante, falta de aptitudes digitales y entendimiento (Arese, 2016).

#### **2.1.4. COVID-19 ante la educación y la inclusión digital.**

Al 20 de marzo de 2020, la Argentina, Bolivia, Chile, Colombia, Ecuador, El Salvador, Guatemala, Honduras, Jamaica, Panamá, el Paraguay, el Perú, Santa Lucía, Trinidad, Tabago, Uruguay y Venezuela habían suspendido las clases en todos los niveles educativos por motivos de pandemia por COVID-19. En varios países se habían aplicado cierres de escuelas. Aunque se han hecho planes para promover el uso de dispositivos digitales en los sistemas educativos, muchas instituciones educativas no cuentan con la infraestructura de tecnologías digitales necesaria. Además, existen brechas en el acceso a las computadoras y a Internet en los hogares.

Los procesos de enseñanza y aprendizaje a distancia no están garantizados. Además, existen diferencias de acceso a los dispositivos digitales y a Internet de banda ancha entre las poblaciones urbanas y rurales, entre las poblaciones que hablan o no el idioma oficial (español o inglés), y entre las poblaciones con y sin discapacidades. América Latina se enfrenta a desafíos en la formación de los docentes en materia de TIC (CEPAL, 2020).

#### **2.1.5. Brecha e inclusión digital.**

Las Naciones Unidas (2001) indican que la existencia de la brecha digital es una consecuencia de la disparidad de los ingresos de cada persona, cuya reducción es un objetivo mundial. Se indica, también, que las TIC son un factor de desarrollo humano y que todos los países deben contar con políticas nacionales que impulsen el uso, difusión y aprovechamiento en diferentes ámbitos y colectivos.

Se identifican cuatro elementos relacionados con la brecha digital:

- Disponibilidad de los equipos que permitan conectarse a Internet.
- Posibilidad de acceder a Internet.
- Conocimiento que permita acceder y navegar en la red.
- Capacidad para convertir la información de la red en conocimiento para el beneficio del usuario.

En diversos estudios se manifiesta que las tecnologías traen grandes beneficios para los que las tienen y saben usarlas, generando la exclusión de aquellos que no las tienen o no las saben usar. Para superar esta situación se planteó un concepto que va más allá del equipamiento, denominado inclusión digital y se concibe “como una forma de inserción social imprescindible para el crecimiento comunitario, que surge como superación de la dimensión netamente cuantitativa y tecnológica del concepto de brecha digital” (Lacruz, 2010).

### **2.1.6. Inclusión digital en Bolivia.**

Bolivia viene dando importantes pasos para lograr la inclusión digital y el impulso a la sociedad de la información. Un claro ejemplo de esto es la creación en 2002 de la Agencia para el Desarrollo de la Sociedad de Información en Bolivia (ADSIB), cuyo objetivo es la reducción de la brecha digital. A principios del año 2004, se crea la Estrategia Boliviana de Tecnologías de la Información y la Comunicación para el Desarrollo (ETIC). A partir de 2007, se establece el Plan Nacional de Inclusión Digital que procura desarrollar la sociedad del conocimiento en el Estado Plurinacional de Bolivia (Sulma Farfán Sossa, 2015).

Respecto del período 2012-2014, se establece la Agenda Digital Boliviana para la implementación de una estrategia de desarrollo de las TIC. En paralelo al establecimiento de políticas sobre las TIC, se ha trabajado en la democratización del acceso a la electricidad como indicador imprescindible para la inclusión digital y el desarrollo de la sociedad de la información. Según el informe del Ministerio de Hidrocarburos y Energía (2010), el acceso a la electricidad es del 90,4% en el área urbana y del 50,8% en el área rural, con una proyección del 100% de acceso en el área urbana para el año 2015 y el año 2025 en el área rural. La ampliación de la cobertura eléctrica, sumada a las tecnologías móviles, está permitiendo que cada vez más usuarios bolivianos tengan acceso a las TIC y sus servicios (Sulma Farfán Sossa, 2015).

Según los datos presentados por el Observatorio Regional de Banda Ancha de la CEPAL, el ingreso de internet en el Estado Plurinacional de Bolivia llega al 30% de la población. El costo de conexión a la banda ancha fija representa el 31,42% del ingreso per cápita y a la banda ancha móvil un 11,28%. También el ancho de banda de internet es la más baja de América del Sur. En respuesta a esta situación, el Gobierno boliviano ha puesto en marcha el satélite Túpac Katari, cuyo objetivo es mejorar el acceso y reducir los costes referidos a Internet, telefonía y televisión. Por otra parte, en la Ley de Educación “Avelino Siñani-Elizardo Pérez” de 2010 se plantea entre los objetivos: “Desarrollar una formación científica, técnica, tecnológica y productiva en complementariedad con los avances de la ciencia y la tecnología universal en todo el Sistema Educativo”, así como: “Promover la investigación científica, técnica, tecnológica y pedagógica en todo el Sistema Educativo Plurinacional” (Sulma Farfán Sossa, 2015).

En este marco se llevan adelante diversos proyectos y programas, tales como: “Una Computadora por Docente”, “Telecentros educativos comunitarios”, el “Portal Educativo Nacional”, “Una Computadora por Estudiante”, desarrollo de material educativo para radio y televisión y otros, encaminados a incorporar a las TIC en todos los niveles de la educación (Sulma Farfán Sossa, 2015).

El proyecto “Una Computadora por Docente” es una de las acciones más importantes para la inclusión digital de los docentes bolivianos. Su objetivo es brindar a los profesores acceso a las TIC para que estos desarrollos procesos educativos mediante las herramientas y los contenidos instalados en las computadoras. En este proyecto se incluye la formación técnica para el uso de los equipos, y pedagógica para el trabajo en la clase. Además de las iniciativas públicas, existen acciones de la cooperación internacional y de organizaciones no gubernamentales nacionales y extranjeras dirigidas a la inclusión

digital en la educación, que ofrecen oportunidades de formación en las TIC, desarrollo de contenidos, difusión de información y equipamiento, entre otros (Sulma Farfán Sossa, 2015).

## **2.2. Machine Learning.**

Actualmente podemos ver que hemos entrado en una nueva era de información en la que las compañías conocen de antemano nuestras preferencias y de acuerdo con nuestro comportamiento en la red, pueden predecir qué productos preferiremos antes de su lanzamiento. Usan nuestra información para generar nuevas campañas de marketing con mayor seguridad de que los productos tendrán una mayor aceptación y al mismo tiempo reducirán los riesgos. Todo este conocimiento es proporcionado gracias a la Ciencia de Datos mediante la técnica del Machine Learning (Aprendizaje Automático). Trabajan con cantidades gigantescas de información, las que con seguridad establecerán un comportamiento en los datos con el tiempo proporcionando patrones que con mucha probabilidad se seguirán dando en el futuro. Es así como se genera la predicción de los datos (Judith, 2018).

Con la cantidad de información que es generada día con día en Internet, ya sea por redes sociales, transacciones comerciales, datos emitidos por distintos dispositivos, etc. existen procesos que aprovechan toda esa información y en lugar de conservarla como data almacenada que solo está ocupando mucho espacio en los servidores, se sigue utilizando para hacer análisis de comportamientos y de algún modo identificar tendencias futuras. Muchas veces se reúne tanta información, que es posible conocer con anticipación y de forma segura cuál será el comportamiento de un grupo de personas o equipos electrónicos en un futuro (Judith, 2018).

### **2.2.1. Inteligencia Artificial.**

Los dispositivos que cuentan con inteligencia artificial pueden ejecutar distintos procesos similares al comportamiento humano, como la devolución de una respuesta por cada entrada (similar a los reflejos de los seres vivos), la búsqueda de un estado entre todos los posibles según una acción o la resolución de problema mediante una lógica formal. Cuando se otorga a estos dispositivos la habilidad de aprender y de discernir, se les convierte en entidades que rozan las capacidades de un superhombre, dado que alcanzan velocidades de procesamiento imposibles para los humanos y no necesitan descansar para funcionar, entre otras ventajas que los ubican por sobre los seres vivos (Judith, 2018).

### **2.2.2. Exploración de Machine Learning**

Machine Learning es una rama de la Inteligencia Artificial (IA) que se encarga de generar algoritmos que tienen la capacidad de aprender y no tener que programarlos de manera explícita. El desarrollador no tendrá que sentarse a programar por horas tomando en cuenta todos los escenarios posibles ni todas las excepciones posibles. Lo único que hay que hacer es alimentar el algoritmo con un volumen gigantesco de datos para que el algoritmo aprenda y sepa qué hacer en cada uno de estos casos. Hay dos tipos de aprendizajes: el supervisado y el no supervisado (Judith, 2018).

### **2.2.3. Aprendizaje Supervisado.**

Es cuando entrenamos un algoritmo de Machine Learning dándole las preguntas (características) y las respuestas (etiquetas). Así en un futuro el algoritmo pueda hacer una predicción conociendo las características. En este tipo de aprendizaje hay dos algoritmos: el de clasificación y el de regresión (Judith, 2018).

#### **2.2.3.1. Algoritmo de clasificación.**

Esperamos que el algoritmo nos diga a qué grupo pertenece el elemento en estudio. Encuentra patrones en los datos que se le otorga y los clasifica en grupos. Luego compara los nuevos datos y los ubica en uno de los grupos y es así como puede predecir de que se trata. La variable por predecir es un conjunto de estados discretos o categóricos. Pueden ser:

- Binaria: {Sí, No}, {Azul, Rojo}, {Fuga, No Fuga}, etc.
- Múltiple: Comprará {Producto1, Producto 2...}, etc.
- Ordenada: Riesgo {Bajo, Medio, Alto}, etc.

#### **2.2.3.2. Algoritmo de regresión.**

En este método lo que se espera es un número. No lo ubica en un grupo, sino que devuelve un valor específico. Por ejemplo, el precio de una casa. El algoritmo tiene el precio de diferentes casas, pequeñas, grandes, en el campo, en la ciudad, etc. y por medio de un gráfico de dispersión, puede predecir el precio correcto de una casa en consulta (Judith, 2018).

### **2.2.4. Aprendizaje No Supervisado.**

Aquí solo le damos las características al algoritmo, nunca las etiquetas. Queremos que nos agrupe los datos que le dimos según sus características. El algoritmo solo sabe que como los datos comparten ciertas características, de esa forma asume que pueda que pertenezcan al mismo grupo (Judith, 2018).

## **2.3. Modelos de Machine Learning**

Los algoritmos de Machine Learning son varios, pero se pueden agrupar en tres modelos:

### **2.3.1. Modelos Lineales.**

Estos tratan de encontrar una línea que se “ajuste” bien a la nube de puntos que se disponen. Aquí destacan desde modelos muy conocidos y usados como la regresión lineal (también conocida como la regresión de mínimos cuadrados), la logística (adaptación de la lineal a problemas de clasificación -cuando son variables discretas o categóricas-). Estos dos modelos tienen el problema del “overfit”, esto significa

que se ajustan “demasiado” a los datos disponibles, con el riesgo que esto tiene para nuevos datos que pudieran llegar. Al ser modelos relativamente simples, no ofrecen resultados muy buenos para comportamientos más complicados (Judith, 2018).

### **2.3.2. Modelos de árbol.**

Son modelos precisos, estables y más sencillos de interpretar básicamente porque construyen unas reglas de decisión que se pueden representar como un árbol. A diferencia de los modelos lineales, pueden representar relaciones no lineales para resolver problemas. En estos modelos, destacan los árboles de decisión y los Random Forrest (una media de árboles de decisión). Al ser más precisos y elaborados, obviamente ganamos en capacidad predictiva, pero perdemos en rendimiento (Judith, 2018).

### **2.3.3. Redes Neuronales.**

Las redes artificiales de neuronas tratan, en cierto modo de replicar el comportamiento del cerebro humano, donde tenemos millones de neuronas que se interconectan en red para enviarse mensajes unas a otras. Esta réplica del funcionamiento del cerebro humano es uno de los “modelos de moda” por las habilidades cognitivas de razonamiento que adquieren. El reconocimiento de imágenes o videos, por ejemplo, es un mecanismo complejo y una red neuronal es lo mejor para realizarlo. El problema, como ocurre con el cerebro humano, es que son lentas de entrenar y necesitan mucha capacidad de cómputo. Quizás sea uno de los modelos que más ha ganado con la “revolución de los datos” (Judith, 2018).

## **2.4. Fases de desarrollo del modelo de machine learning.**

### **2.4.1. Fase de entrenamiento.**

En esta fase se tiene una cantidad enorme de datos, de la cual se separa una parte para entrenar al algoritmo y darle toda esta información para que encuentre los patrones necesarios y después pueda hacer predicciones (Judith, 2018).

### **2.4.2. Fase de prueba.**

El resto de los datos que quedan, se van a usar para hacer las pruebas. Así le podemos hacer preguntas al algoritmo y evaluar si las respuestas están bien o mal, y saber si está aprendiendo o no. Si vemos que no coinciden los datos, tendremos que agregar más datos o cambiar el método que estamos utilizando. Pero si se observa que hay entre un 80% a 90% de respuestas correctas, podemos decir que hay un buen grado de aprendizaje y poder utilizar ese algoritmo (Judith, 2018).

### **2.4.3. Fase de evaluación.**

Las métricas de rendimiento juegan un papel muy importante en la evaluación de los modelos de problemas de clasificación, donde se busca discriminar diferentes algoritmos Machine y Deep Learning, con la finalidad de facilitar la elección del mejor algoritmo dependiendo del objetivo de investigación.

Las métricas son: Matriz de confusión, Accuracy, Recall, F1-score, curva ROC, AUC (Borja-Robalino, 2020).

#### **2.4.3.1. Matriz de confusión.**

es una técnica para resumir el rendimiento de un algoritmo de clasificación. Consta de una tabla de dos dimensiones: real y predicción. Las columnas de la matriz indican la clase observada o real, y las filas indican la clase predicha. La dimensión real se refiere a los datos reales, los cuales van a ser comparados con las predicciones. ver la figura 2-1 (Bernedo, 2021).

		<b>Clasificador 1</b>	
		<b>Clase 1: Positivo</b>	<b>Clase 2: Negativo</b>
<b>Clasificador 2</b>	<i>Clase 1: Positivo</i>	<i>f11 = Verdadero positivo</i>	<i>f10 = Falso negativo</i>
	<i>Clase 2: Negativo</i>	<i>f01 = Falso positivo</i>	<i>f00 = Verdadero negativo</i>

**Figura 2-1: Matriz de confusión.**

**Fuente:** (Borja-Robalino, 2020).

#### **2.4.3.2. La Accuracy.**

Es la métrica más utilizada por investigadores en el caso de dos clases y multiclas, debido a su facilidad de cálculo y comprensión para evaluar la efectividad general del algoritmo. Sin embargo, una de sus principales desventajas es que produce menos valores discriminatorios y distintivos para el caso multiclas desbalanceado. Se refiere a lo cerca que está el resultado de una medición al valor verdadero (Bernedo, 2021).

#### **2.4.3.3. Recall o sensibilidad.**

Es una medida que permite conocer la proporción de casos positivos que fueron correctamente clasificados. En un modelo perfecto el Recall es igual a 1 para cada clase. Desde el punto de vista analítico un investigador busca aumentar el Recall sin afectar el valor de la Accuracy (Borja-Robalino, 2020).

#### **2.4.3.4. F1-score.**

Fusiona las métricas Accuracy con Recall, presentando diferencias en el rendimiento de un clasificador que no son revelados únicamente con la Accuracy. Es directamente proporcional al aumento de las dos medidas, por lo tanto, valores altos de F1-score demuestran que el algoritmo de clasificación predice de mejor manera la clase positiva (Borja-Robalino, 2020).

Otra métrica utilizada a menudo para la evaluación de rendimiento en clasificadores de dos clases es la curva ROC, que se trata de una gráfica bidimensional de la sensibilidad versus  $(1 - \text{especificidad})$  por cada clase. La medida de comparación en esta grafica es la AUC que corresponde al área bajo la curva

ROC y sus valores se encuentran entre 0 y 1, considerando que en el caso totalmente aleatorio se tiene una AUC igual a 0,5 (Borja-Robalino, 2020).

## **2.5. Ámbitos de la aplicación de algoritmos machine learning.**

Productos que utilizan algoritmos de Machine Learning:

Vehículos no tripulados que se conducen solos, brazo robótico que juega ajedrez, reconocimiento facial de Facebook para identificar contactos, Microsoft Cortana, asistente personal inteligente para diferentes dispositivos, motores de búsqueda que ofrecen información de acuerdo a las preferencias de los usuarios, Machine Translation usado por el traductor de Google, que reconoce palabras en más de 100 idiomas humanos, Google Trends, son las tendencias de búsquedas en Google, Google N Gram Viewer, indexa libros que tiene Google escaneados y sus términos gramaticales, Siri, que convierte conversaciones habladas a texto (STT – Speech To Text).

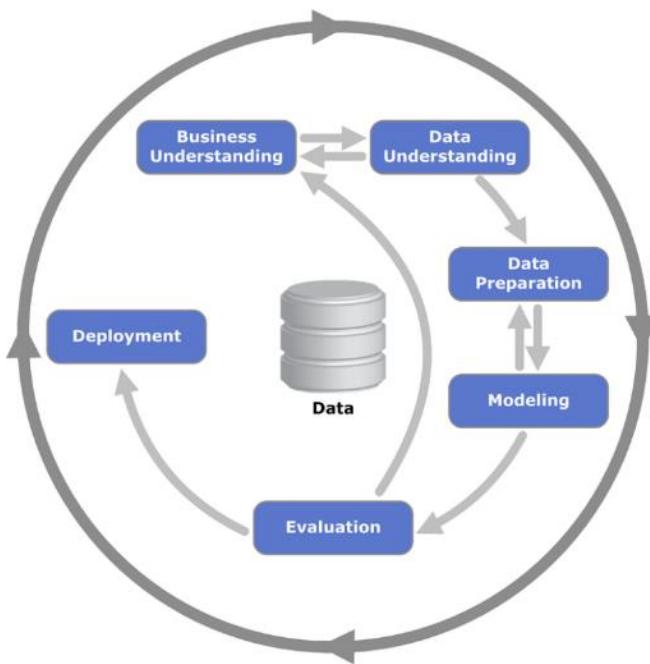
Procesos que hacen uso de Machine Learning:

Detectar fraudes en transacciones bancarias, detectar intrusiones en una red de comunicaciones de datos, predecir fallos en equipos tecnológicos, prever qué proyectos serán más rentables el próximo año y con un menor riesgo, seleccionar clientes potenciales basándose en comportamientos en las redes sociales, interacciones en la web, etc; predecir el tráfico urbano y dar rutas alternativas, conocer anticipadamente qué partido político ganará las próximas elecciones analizando los comentarios de los usuarios en las redes sociales, prevenir la deserción de clientes en una empresa de telefonía, predecir las ventas de los años siguientes analizando comportamiento actual de los clientes, conocer las preferencias de los clientes a través de sus operaciones en la red, hacer prediagnósticos médicos basados en síntomas del paciente, cambiar el comportamiento de una App móvil para adaptarse a las costumbres y necesidades de cada usuario.

## **2.6. Metodología CRISP-DM.**

CRISP-DM (Cross Industry Standard Process for Data Mining), es un modelo de proceso de minería de datos que describe una manera en la que los expertos en esta materia abordan el problema.

CRISP-DM incluye un modelo y una guía, estructurados en seis fases, algunas de las cuales son bidireccionales, es decir que de una fase en concreto se puede volver a una fase anterior para poder revisarla, por lo que la sucesión de fases no tiene por qué ser ordenada desde la primera hasta la última. En la figura 2-2 se puede observar las fases en las que se divide CRISP-DM y las posibles secuencias a seguir entre ellas, (Cortina, 2015)



**Figura 2-2: Fases de la metodología CRISP-DM.**

Fuente: [healthdataminer.com/data-mining/crisp-dm-una-metodologia-para-mineria-de-datos-en-salud/](http://healthdataminer.com/data-mining/crisp-dm-una-metodologia-para-mineria-de-datos-en-salud/)

### **2.6.1 Comprensión del negocio.**

Esta primera fase es probablemente la más importante y une las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva de negocio, con el fin de convertirlos en objetivos técnicos y en un plan de proyecto. Sin lograr comprender dichos objetivos, ningún algoritmo por muy sofisticado que sea, permitirá obtener resultados fiables. Para obtener el mejor provecho de la minería de datos, es necesario entender de la manera más completa el problema que se desea resolver, esto permitirá recolectar los datos correctos e interpretar correctamente los resultados. En esta fase, es muy importante la capacidad de poder convertir el conocimiento adquirido del negocio en un problema de minería de datos y en un plan preliminar cuya meta sea el alcanzar los objetivos del negocio, (Cortina, 2015).

### **2.6.2 Comprensión de los datos.**

Esta segunda fase comprende la recolección inicial de los datos con el objetivo de establecer un primer contacto con el problema, familiarizarse con ellos, identificar su calidad y establecer las relaciones más evidentes que permitan definir las primeras hipótesis. Esta fase junto a las dos siguientes fases es las que demandan el mayor esfuerzo y tiempo en un proyecto de minería de datos. Por lo general si la organización cuenta con una base de datos corporativa, es deseable crear una nueva base de datos específica para el proyecto de DM (Data Mining), ya que durante el desarrollo del proyecto es posible que se generen frecuentes y abundantes accesos a la base de datos con el fin de realizar consultas y probablemente se produzcan modificaciones, lo cual podría generar muchos problemas, (Cortina, 2015).

### **2.6.3 Preparación de los datos.**

En esta fase y una vez efectuada la recolección inicial de los datos, se procede a su preparación para adaptarlos a las técnicas de minería de datos que se van a utilizar posteriormente, éstas pueden ser técnicas de visualización de datos, de búsqueda de relaciones entre variables u otras medidas para explotación de los datos. La preparación de los datos incluye las tareas generales de selección de datos a los que se va a aplicar una determinada técnica de modelado, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato. Esta fase se encuentra relacionada con la fase de modelado, ya que, en función de la técnica de modelado elegida, los datos requieren ser procesados de una manera o de otra, por esta razón las fases de preparación y de modelado interactúan de forma permanente, (Cortina, 2015).

#### **2.6.3.1. Conversión One Hot Encoding.**

One-Hot Encoding es una técnica utilizada en el procesamiento de datos, especialmente en el ámbito de la inteligencia artificial y el aprendizaje automático, para representar variables categóricas como vectores binarios. En lugar de utilizar una única columna para representar las categorías, se crea una columna binaria separada para cada categoría posible. Estas columnas binarias se llaman "indicadores" o "dummy variables".

El proceso implica asignar el valor 1 a la columna correspondiente a la categoría presente y 0 a todas las demás columnas. Esto crea una representación numérica que es más adecuada para muchos algoritmos de aprendizaje automático, que generalmente operan mejor con datos numéricos.

Si se tiene una columna "Color" con categorías "Rojo", "Verde" y "Azul", el One-Hot Encoding convertiría esta información en tres columnas: "Rojo" (con valores binarios 1 o 0), "Verde" (con valores binarios 1 o 0), y "Azul" (con valores binarios 1 o 0). Cada fila indicaría el color presente con un 1 y el resto con 0, (Cortina, 2015).

#### **2.6.3.2. Método Z-score.**

Los valores atípicos son los que se encuentran muy alejados del conjunto común de los datos, y suelen causar errores en el momento del modelado, por lo cual se debe eliminar estos valores. Para esto se usa el método z-score.

Z-score indica cuántas desviaciones estándar, está por encima o por debajo de la media. Un Z-score positivo significa que el valor está por encima de la media, mientras que un Z-score negativo indica que está por debajo de la media. Los Z-scores también son útiles para identificar valores atípicos (outliers) en un conjunto de datos, ya que valores con Z-scores extremadamente altos o bajos pueden indicar puntos que se apartan significativamente de la norma en términos de la distribución de los datos.

La elección del valor 3 como umbral para identificar valores atípicos (outliers) con el Z-score es una convención común en estadísticas. Tradicionalmente, se considera que los valores que tienen un Z-score

mayor que 3 o menor que -3 son atípicos, lo que significa que están a más de tres desviaciones estándar de la media.

Esta elección se basa en la suposición de que, en una distribución normal aproximadamente el 99.7% de los datos deberían estar dentro de tres desviaciones estándar de la media. Por lo tanto, cualquier punto de datos que caiga fuera de este rango se considera atípico.

#### **2.6.4 Modelado.**

En esta fase de CRISP-DM se seleccionan las técnicas de modelado más apropiadas para el proyecto de minería de datos específico. Las técnicas a utilizar en esta fase se eligen en función de los siguientes criterios:

- Ser apropiada para el problema.
- Disponer de los datos adecuados.
- Cumplir los requisitos del problema.
- Tiempo adecuado para obtener un modelo.
- Conocimiento de la técnica.

Previamente al modelado de los datos se debe determinar un método de evaluación de los modelos que permita establecer el grado de adecuación de cada uno de ellos. Después de concluir estas tareas genéricas se procede a la generación y evaluación del modelo. Los parámetros utilizados en la generación del modelo dependen de las características de los datos y de las características de precisión que se quieran lograr con el modelo, (Cortina, 2015).

#### **2.6.5 Evaluación.**

En esta fase se evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema. Debe considerarse además que la fiabilidad calculada para el modelo se aplica solamente para los datos sobre los que se realizó el análisis. Es preciso revisar el proceso, teniendo en cuenta los resultados obtenidos, para poder repetir algún paso anterior, en el que se pueda haber cometido algún error. Considerar que se pueden emplear múltiples herramientas para la interpretación de los resultados. Si el modelo generado es válido en función de los criterios de éxito establecidos en la fase anterior, se procede a la explotación del modelo, (Cortina, 2015).

#### **2.6.6 Despliegue o implantación.**

En esta fase, y una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio, esto puede hacerse por ejemplo cuando el analista

recomienda acciones basadas en la observación del modelo y sus resultados, o por ejemplo aplicando el modelo a diferentes conjuntos de datos o como parte del proceso (en análisis de riesgo de créditos, detección de fraudes, etc.). Generalmente un proyecto de minería de datos no concluye en la implantación del modelo, ya que se deben documentar y presentar los resultados de manera comprensible para el usuario con el objetivo de lograr un incremento del conocimiento. Por otra parte, en la fase de explotación se debe asegurar el mantenimiento de la aplicación y la posible difusión de los resultados, (Cortina, 2015).

## 2.7. Google Colaboratory.

En la figura 2-2, se observa como Google Colaboratory también conocido como Google Colab, constituye una excelente opción dado que se ofrece como un servicio gratuito de Google que proporciona una interfaz de Jupyter Notebook que permite compartir el trabajo simplemente otorgando acceso al cuaderno de trabajo, como se hace habitualmente con los archivos de Google Drive; sólo se necesita contar con un navegador Web y conexión a Internet (Esnaola).

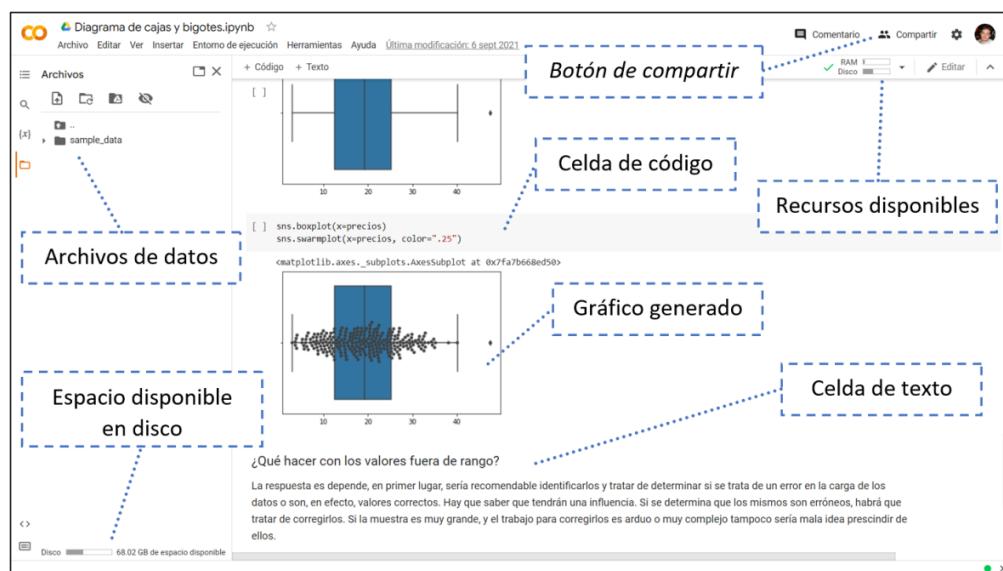


Figura 2-3: Interfaz de Google Colab

Fuente: Google Colaboratory 2023.

### 2.7.1 Características y funcionamiento de Google Colab.

Los cuadernos de trabajo (Jupyter Notebooks) de Google Colab se ejecutan en máquinas virtuales (MVs) basadas en Linux, que son proporcionadas y mantenidas por Google. El cálculo se puede realizar con unidades de procesamiento central, o se lo puede acelerar a través de unidades especializadas de procesamiento gráfico y de procesamiento de tensores. El hardware disponible para cada MV varía según la sesión, pero generalmente incluye una GPU NVIDIA de primera línea (K80, T4 o P100); una CPU

Intel® Xeon® de 2 a 2.3 GHz; alrededor de 8 a 13 GB de RAM; y de 39 a 70 GB de espacio libre en el disco duro de la MV (Esnaola).

Dado que los cuadernos de trabajo de Google Colab están diseñados para uso interactivo, en lugar de experimentos de ejecución prolongada, las MVs se desconectan después de un tiempo de inactividad y limitan las sesiones a 12 horas diarias. Sin embargo, estas características se pueden mejorar suscribiendo a los planes “Pro” o “Pro+” (Google Research, s.f.) que requieren un pago mensual (Esnaola).

las máquinas virtuales basadas en la nube de Google Colab están precargadas con paquetes comunes de inteligencia artificial (numpy, torch, tensorflow, etc.). Y, si alguna biblioteca de código no está precargada en la instalación, como el framework Deep, se lo puede instalar fácilmente ejecutando en una celda del cuaderno de trabajo de Google Colab: “! pip install deep”. Otro punto fuerte de Google Colab es su conexión nativa con el servicio Google Drive, lo que posibilita utilizar el propio espacio de almacenamiento en la nube para guardar los cuadernos de trabajo; los archivos de datos; los gráficos generados, entre otros; y compartirlos fácilmente con otros usuarios administrando los permisos de acceso. Esta característica permite potenciar el proceso de aprendizaje, pues se puede compartir el trabajo en progreso y solicitar asistencia ante un error determinado, ya sea por parte del equipo docente o de colegas (Esnaola).

### **2.7.2 Utilidad de Google Colab.**

Google Colab tienen la posibilidad de guardar una copia en su propio espacio de Google Drive, e intervenirlos de una u otra forma, o ejecutarlos “en línea” sin necesidad de guardar los cambios. Esta característica resulta atractiva para, experimentar “qué pasaría si...”, sin temor a romper nada, lo que permite profundizar la aprehensión de los contenidos. Además, Google Colab constituye el mecanismo recomendado para desarrollar las actividades de programación. Los enunciados de los ejercicios de cada actividad se comparten en un cuaderno de trabajo que acompaña a la consigna. Cada alumna o alumno desarrolla la resolución de los ejercicios presentados en una copia personal del mismo. Finalmente, pueden subir el cuaderno de trabajo resultante a través de una actividad “tarea” de la Plataforma de Educación Digital, o podrían directamente compartir su cuaderno de trabajo (Esnaola).

## **2.8. Tableau.**

Tableau es una plataforma de análisis visual que transforma la manera en que usamos los datos para resolver problemas. Además, permite a las personas y las organizaciones sacar el máximo partido de los datos (Tableau, 2003).

### **2.8.1 Función de Tableau.**

La plataforma de Tableau es la opción de inteligencia de negocios moderna líder en el mercado. Hace que sea más fácil explorar y administrar los datos. Asimismo, permite descubrir y compartir información más rápidamente a fin de generar grandes cambios en los negocios y en el mundo (Tableau, 2003).

Su misión es ayudar a las personas a que puedan ver y comprender los datos. Por ese motivo se diseñó esta plataforma pensando en los usuarios, ya sean analistas, científicos de datos, estudiantes, docentes, ejecutivos o usuarios corporativos, es la plataforma de análisis integral más eficaz, segura y flexible (Tableau, 2003).

### 2.8.2 Tableau como analizador visual.

Tableau tiene como objetivo mejorar el flujo del análisis y poner los datos al alcance de las personas a través de la visualización como se observa en la figura 2-3. Los cofundadores Chris Stolte, Pat Hanrahan y Christian Chabot desarrollaron y patentaron la tecnología central de Tableau: VizQL. Esta tecnología permite expresar datos visualmente al transformar acciones de arrastrar y soltar en consultas de datos, en una interfaz intuitiva (Tableau, 2003).

Se crearon soluciones para ayudar a todas las personas que trabajan con datos a obtener respuestas más rápidamente y descubrir información nueva. Este proceso incluye lograr que el aprendizaje automático, las estadísticas, el lenguaje natural y la preparación de datos inteligente sean más eficaces y permitan aumentar la creatividad del usuario para el análisis. Además de ofrecer una plataforma de análisis integrada y completa, brinda recursos comprobados. Estos ayudan a los clientes a implementar y adaptar una cultura basada en los datos que promueve la resiliencia y el valor a través de resultados eficaces (Tableau, 2003).

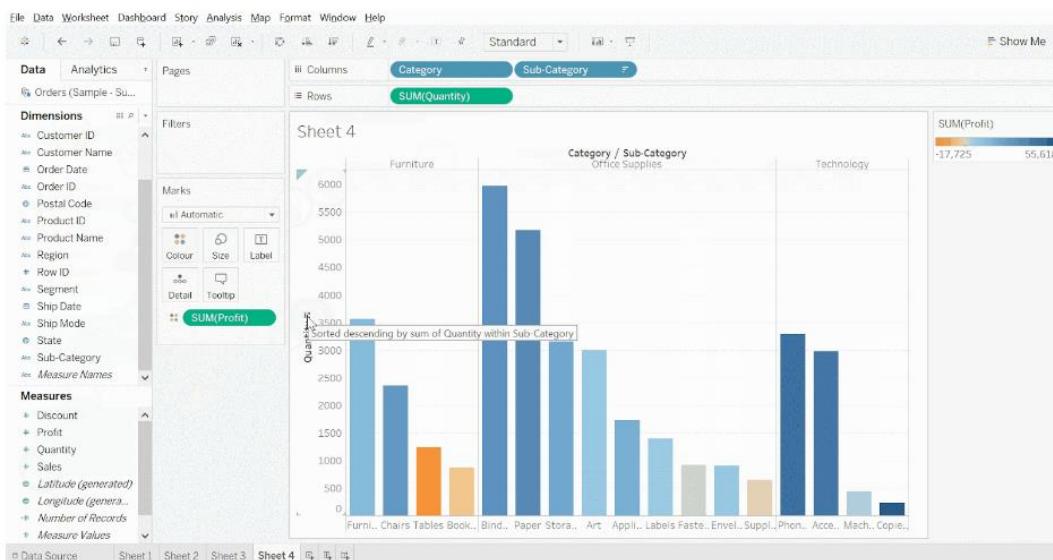


Figura 2-4: Interfaz de Tableau.

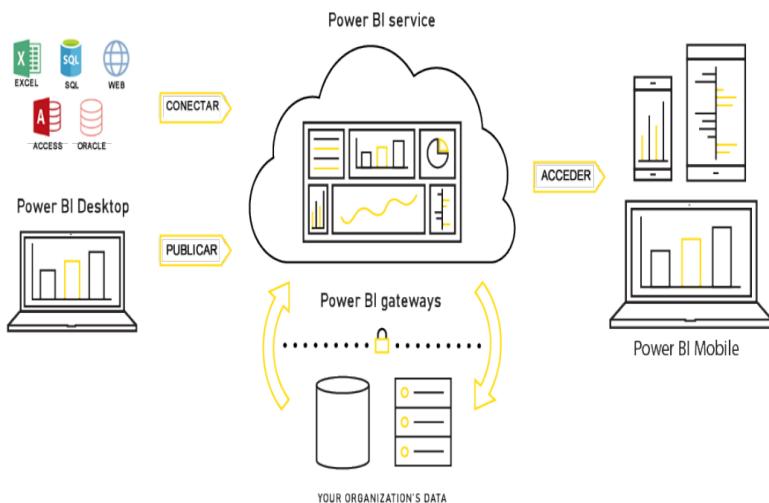
Fuente: [www.simplilearn.com](http://www.simplilearn.com) 2023

### 2.9. Power BI.

Microsoft presentó su idea de Inteligencia de negocios en el 2009, anunciando Power Pivot para Microsoft Excel 2010. Power Pivot fue el primer paso de Microsoft en el campo Inteligencia de negocios

de autoservicio. Power BI es una evolución de los complementos anteriormente disponibles en Excel: Power Pivot, Power Query, Power View Y Power Map.

Proporciona análisis de datos orientados al usuario y capacidades de visualización para mejorar el proceso de toma de decisiones y las ideas de negocio. Power BI es una aplicación de Inteligencia de negocios de autoservicio basada en la nube, lo que significa que puede crear e implementar soluciones rápidamente con datos procedentes de una amplia variedad de fuentes de datos, sistemas y aplicaciones en la nube. Todo esto está alojado en la nube empresarial de confianza de Microsoft. Power BI es una colección de servicios de software, aplicaciones y conectores que funcionan conjuntamente para convertir orígenes de datos sin relación entre sí en información coherente, interactiva y atractiva visualmente. Tanto si se trata de una sencilla hoja de cálculo de Excel como de una colección de almacenes de datos híbridos locales o basados en la nube, permite conectar fácilmente 42 los orígenes de datos, visualizar lo más importante y compartirlo con cualquier usuario, (AYALA, 2017).



**Figura 2-5: Conexión y funcionamiento de Power Bi.**

Fuente: (AYALA, 2017).

## 2.10. Clasificación de impacto de inclusión digital.

Determinar si una escuela tiene un alto o bajo impacto en cuestión de la inclusión digital, resulta una tarea difícil, pero si se cuenta con la información necesaria y adecuada para evaluar si hubo o no un alto o bajo impacto no ve una tarea tan complicada.

Los autores “Carles Sigalés, Josep M. Mominó, Julio Meneses y Antoni Badia” del libro “INTEGRACION DE INTERNET EN LA EDUCACION ESCOLAR ESPAÑOLA” mencionan. Lo que más valora el profesorado, por lo que al acceso a internet se refiere, es la gran variedad de recursos (bases de datos, sitios web, blogs, etc.) interesantes para su asignatura que habitualmente puede encontrar. El 72,5% del profesorado se manifiesta en este sentido. En cuanto a los recursos tecnológicos, el

porcentaje de profesores que piensa que la disponibilidad de ordenadores conectados a internet en su centro docente facilita un uso habitual de estas tecnologías es también mayoritario, pero en una proporción inferior (58,4%). un porcentaje similar (el 57,6%) considera suficiente la velocidad y la calidad de la conexión a internet con la que puede trabajar con sus alumnos; y un 54% opina que puede acceder con facilidad a programas y a otras aplicaciones informáticas de utilidad para sus asignaturas. por segmentos, los profesores de los centros privados y los de secundaria son los que ven más dificultades para usar con frecuencia las TIC, a causa del número de ordenadores conectados a internet de que disponen en su centro, y por no tener un fácil acceso a programas y aplicaciones informáticas útiles para sus asignaturas. En cambio, los menos satisfechos con la calidad y la velocidad de la conexión a internet son los profesores de los centros públicos, (Carles Sigalés, 2009).

Por lo que al acceso a internet y a la disponibilidad de recursos se refiere, hay que destacar que, mientras los directores ven en los ordenadores disponibles en el centro el principal incentivo para utilizar las TIC, los profesores valoran básicamente los recursos de información relacionados con su asignatura que les proporciona internet. La mayoría de los directores y profesores ven en las TIC un instrumento para mejorar la consecución de los objetivos educativos y la calidad de los aprendizajes de sus alumnos, (Carles Sigalés, 2009).

Wallace (2004) plantea que las TIC y especialmente internet modifican en mayor o menor grado los límites, la autoridad, la estabilidad, el contexto pedagógico y el contexto disciplinario en las prácticas educativas. El propio Castells (2004) destaca que las TIC, a causa de la velocidad con la que procesan la información y permiten la comunicación y de su infinita capacidad de reconfiguración, revolucionan la velocidad y la extensión con la que se pueden propagar las innovaciones. La velocidad que las TIC imprimen a las innovaciones choca, a menudo, con el ritmo mucho más lento que suelen tener este tipo de cambios en las instituciones escolares. El profesorado de la educación obligatoria, con respecto a los conocimientos y habilidades digitales que, resulta ser también bastante competente, tanto en el dominio técnico de los ordenadores como en el uso de internet, (Carles Sigalés, 2009).

En este sentido, en el extremo superior de su valoración, al menos dos tercios del profesorado considera que estaría capacitado para localizar en internet recursos útiles y fiables para su asignatura (82%), preparar las clases de manera que los alumnos tengan que utilizar las TIC (64,7%), saber cuáles son las situaciones de enseñanza y aprendizaje más apropiadas para utilizar las TIC (61,4%), o utilizarlas para hacer presentaciones o explicaciones más efectivas (60,3%), (Carles Sigalés, 2009).

El autor Joel Abelardo Vilca Gallardo, en su monografía “EFICIENCIA EN LA INCLUSION DE TICs EN EL SISTEMA EDUCATIVO EN BOLIVIA UTILIZANDO ANALITICA DE DATOS”, afirma que los mayores problemas que afronta la población boliviana es la falta de alfabetización digital y los recursos necesarios para lograr un óptimo rendimiento de aprendizaje; la falta de apoyo político también es otro factor por la cual la educación en Bolivia aún se encuentra con brechas digitales. El autor denota que las características más importantes son la alfabetización en TIC y la accesibilidad a recursos tecnológicos.

Según los autores del anterior libro mencionado, las características más sobresalientes y las que catalogan a una buena escuela son aquellas que cuentan con internet, aulas tecnológicas, computadoras tanto para maestro y estudiantes, implementación de TIC en aula, conocimientos ofimáticos e informáticos por parte de los docentes, tal como se observa en Tabla 2-1.

**Tabla 2-1: Características que catalogan una buena escuela con inclusión digital.**

Característica 1	Característica 2	Característica 3	Característica 4	Característica 5
Contar con internet en la escuela.	Contar con aula tecnológica.	Contar con computadora tanto para profesor y estudiante.	Implementar TIC en aulas.	Conocimiento ofimático e informático por parte de los profesores.

**Fuente:** Elaboración propia según libro “Integración de internet en la educación escolar española”, 2023.

Con la ayuda de 24 profesores del departamento de Cochabamba que fueron participes de una encuesta en la cual se recaudó información de, si cuentan con acceso a internet, si tiene un aula tecnológica con computadoras, que método de enseñanza usan, si permiten que los estudiantes usen los celulares en aula, si se implementan las TIC, el nivel de conocimiento que tienen los maestros en ofimática e informática y si el uso de Tecnología mejoró la educación en sus unidades educativas, tal como se observa en la Tabla 2-2.

**Tabla 2-2: Tabla con variables fundamentales de las encuestas.**

Internet en colegio	Aula computacional en colegio	Requerimiento del colegio	Método de enseñanza	uso de celular en aula	Conocimiento en TIC	Frecuencia de uso de TIC	Uso de TIC mejora el aprendizaje de estudiantes	Conocimientos de ofimática	Conocimientos de informática
Si	Si	Herramientas tecnológicas	Práctico	Si	Si	Siempre	Si	Bueno	Bueno
No	Si	Herramientas tecnológicas	Práctico	No	Si	Siempre	Si	Bueno	Regular
Si	Si	Biblioteca Digital	Teórico	Si	Si	Siempre	Si	Bueno	Regular
Si	Si	Herramientas tecnológicas	Teórico	Si	Si	Ocasionalmente	Si	Bueno	Regular
No	Si	Computadora para estudiantes	Teórico	No	Si	Rara vez	Si	Regular	Bueno
Si	Si	Aula computacional	Práctico	No	Si	Ocasionalmente	No	Bueno	Regular
No	No	Herramientas tecnológicas	Teórico	No	Si	Siempre	Si	Bueno	Bueno
Si	Si	Aula computacional	Práctico	Si	Si	Siempre	Si	Bueno	Bueno
Si	Si	Internet en aulas	Teórico	Si	Si	Siempre	Si	Bueno	Regular
Si	Si	Herramientas tecnológicas	Práctico	Si	No	Ocasionalmente	Si	Regular	Regular
Si	Si	Internet en aulas	Teórico	Si	Si	Ocasionalmente	Si	Bueno	Bueno
No	Si	Herramientas tecnológicas	Práctico	Si	Si	Siempre	Si	Bueno	Bueno
Si	Si	Internet en aulas	Teórico	No	Si	Siempre	Si	Bueno	Regular
Si	Si	Biblioteca Digital	Práctico	No	Si	Siempre	Si	Bueno	Bueno
No	No	Internet en aulas	Práctico	No	Si	Ocasionalmente	Si	Regular	Regular

**Fuente:** Elaboración propia, 2023.

Según las características mencionadas en la Tabla 2-1, del libro “Integración de internet en la educación escolar española”, y los proyectos de los otros 2 autores que mencionan que son indicadores que ayudan a verificar si una escuela tiene un alto o bajo impacto en inclusión digital. Basados en lo anterior mencionado se puede decir que para que una escuela tenga un alto impacto en inclusión digital debe contar con acceso a internet, contar con un aula tecnológica donde cuente con computadoras, que sus métodos de enseñanza sean diferentes a la manera clásica que era solo teoría, que los estudiantes puedan interactuar de manera digital con los maestros por medio del uso de celulares en clases, que los maestros

tengan conocimientos y usen las TIC en sus clases y que los maestros tengan conocimientos en ofimática e informática. Donde aquellos colegios que no cuenten con estas características serán considerados con impacto de inclusión digital bajo, tal como se observa en la Tabla 2-3.

**Tabla 2-3: Valores calificativos para determinar alto y bajo impacto.**

Tipo de impacto	Internet en colegio	Aula computacional en colegio	Método de enseñanza	Uso de celular en aula	Conocimiento en TIC	Frecuencia de uso de TIC	Uso de TIC mejoró el aprendizaje de estudiantes	Conocimientos de ofimática	Conocimientos de informática
Alto	Si	Si	Práctico	Si	Si	Siempre, Ocasionalmente, Rara vez	Si	Bueno, Regular	Bueno, Regular
Bajo	No	No	Teórico	No	No	Nunca	No	Malo	Malo

**Fuente:** Elaboración propia, 2023.

### 3. Metodología

El gobierno boliviano ha invertido en la educación pública, lo que hace que sea un lugar ideal para estudiar el impacto de la inclusión digital en la educación. Como objetivo de estudio se tiene clasificar el impacto de la inclusión digital en las escuelas públicas en función a la infraestructura y enseñanza en base a la encuesta recolectada por AGETIC. La inclusión digital puede tener un impacto alto en el aprendizaje de los estudiantes. Por ejemplo, la tecnología puede ayudar a los estudiantes a aprender a su propio ritmo, a Google Colaborar con otros estudiantes y a acceder a una gama más amplia de recursos educativos.

#### 3.1. Área de estudio

Este proyecto se centra en todo el territorio boliviano, Bolivia es un país ubicado en la parte central de América del Sur. Es un país diverso, con una población de aproximadamente 11 millones de personas. Es un país con una mezcla de culturas indígenas, europeas y africanas, tiene un clima variado que va desde el clima tropical de la región amazónica hasta el clima frío de la región andina, también tiene una variedad de paisajes que incluyen selvas tropicales, montañas, lagos y desiertos.

Es un lugar interesante para estudiar el impacto de la inclusión digital en las escuelas públicas. El país tiene una población diversa, con una gran cantidad de escuelas públicas en diferentes condiciones socioeconómicas y geográficas. Esto permite estudiar el impacto de la inclusión digital en un entorno diverso.

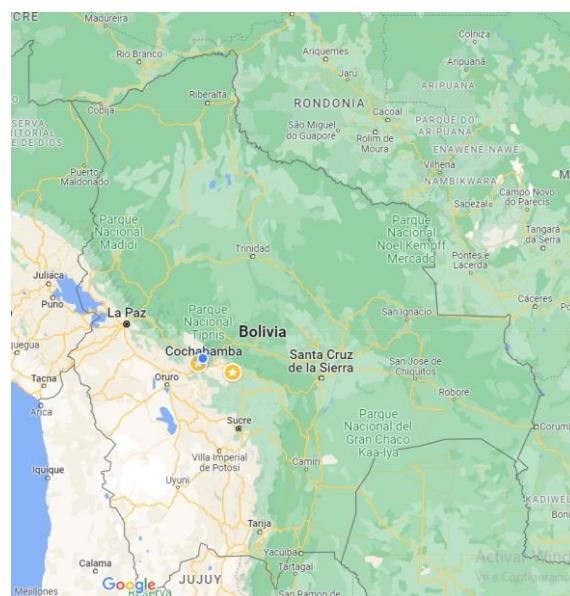


Figura 3-1: Mapa de Bolivia.

Fuente: Google Maps, 2023.

### 3.2. Flujograma metodológico

En la figura 3-2 el siguiente diagrama nos describe todo el proceso que contempla el estudio para la clasificación del impacto de la inclusión digital en las escuelas públicas en base a su infraestructura y enseñanza.

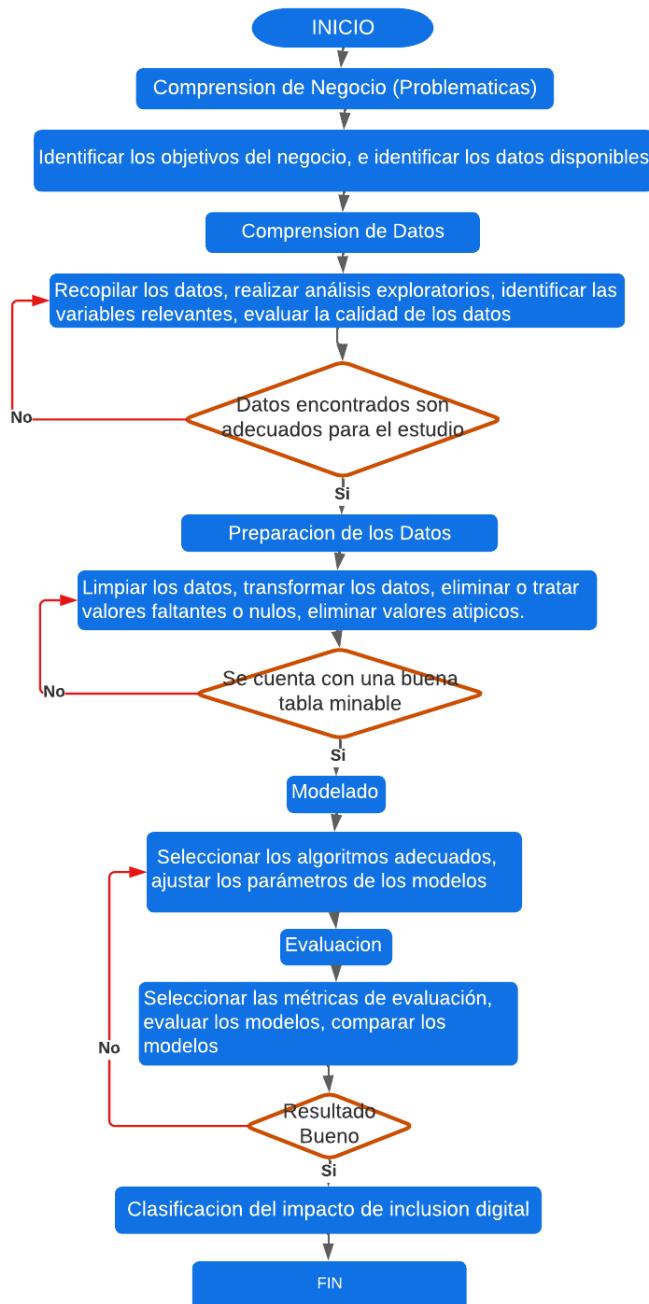


Figura 3-2: Flujograma metodológico.

Fuente: Elaboración propia en Lucid app, 2023.

### **3.2.1. Etapas de desarrollo.**

A continuación, se presentan los siguientes pasos para realizar todo el proceso de desarrollo:

- **Comprensión del Negocio (Problemáticas).**

Como partida inicial se identifica los objetivos a los que el proyecto apunta llegar, también se identifica que información será necesaria para realizar dicho proyecto.

- **Comprensión de datos.**

Una vez ya conocidos los objetivos y que información que es necesaria, se comienza a recopilar el conjunto de datos que ayudara a alcanzar el objetivo. Se elige las variables relevantes y se comienza a realizar el Análisis Exploratorio de Datos (AED).

- **Preparación de los datos.**

Es la etapa donde ya se comienza a preparar los datos para la creación del modelo de machine learning, donde se limpian los datos, se eliminan variables repetidas, se hace tratamiento de valores nulos, tratamiento de valores atípicos y una transformación de valores cualitativos a cuantitativos con la conversión de One Hot Encoding, dejando así listo para que modelo utilice el conjunto de datos ya preparado.

- **Modelado.**

En esta etapa se hace la selección de los algoritmos (árboles de decisión, máquinas de vectores de soporte (SVM), clasificación de Naive Bayes, K vecinos más cercanos (KNN), regresión logística y redes neuronales) a utilizar para la creación de los modelos y así poder comparar sus métricas y ver cual logra destacar más. Para seguidamente en base a la variable objetivo se separa las variables predictoras X que llegarían a ser las variables que no son objetivos, y la variable objetivo Y, luego se divide los datos en conjuntos de entrenamiento y prueba, siendo el 80% de los datos para el entrenamiento y el 20% para la prueba, seguidamente se crea el modelo, para luego pasar a su respectivo entrenamiento,

- **Evaluación.**

Con las métricas como Accuracy, precisión, Recall, f1\_score, y roc\_auc, se procede a evaluar los modelos entrenados dependiendo el valor de estas métricas, aquellos que estén mas cerca del valor 1, son más que tendrán más exactitud en sus predicciones. También una forma de complementar los resultados es usar la matriz de confusión.

### 3.2.2. Clasificación de impacto

El modelo se utiliza para clasificar el impacto de la inclusión digital en las escuelas públicas. Se aplica a un conjunto de datos que incluye información sobre la infraestructura (internet, aula tecnológica, computadoras para estudiantes y profesores, uso de celulares) y enseñanza (conocimientos en TIC, implementación de TIC, conocimientos en ofimática e informática, uso de herramientas de Google como Google Drive, Google Académico, correo electrónico) de las escuelas públicas. Clasifica el impacto de la inclusión digital en las escuelas públicas en una de las siguientes categorías:

- Alto impacto.
- Bajo impacto.

### 3.3. Fuentes de información

#### 3.3.1. Fuente de información secundaria.

El conjunto de datos de los profesores y sus escuelas públicas se obtuvo, de una página web abierta (datos.gob.bo) llamado “Encuesta Final-Profesores de Inclusión Digital”. Este dataset fue recopilado por AGETIC, mediante encuestas realizada a los profesores de varios colegios de los 9 departamentos de Bolivia en el año 2021, el cual se encuentra en un archivo de formato "Comma-Separated Values" (csv), con 187 columnas o variables y 871 registros. Este cuenta con: datos personales de maestro, sus conocimientos, uso de redes sociales, uso de herramientas, y si las escuelas cuentan con ambientes tecnológicos, que requerimientos son necesarios y un punto muy importante para todo este análisis es si estos cuentan con acceso de Internet, tal y como se puede ver en la tabla 3-1, para tener información más completa. ver Anexo 1.

**Tabla 3-1: Encuesta Final-Profesores de Inclusión Digital.**

Sexo	Edad	Departamento	Ciudad	Colegio	Años de experiencia	Tiempo en el colegio	Años en el Colegio
Masculino	32	Santa Cruz	Santa Cruz de la Sierra	NACIONAL COTOCO	20	Más de un año	15
Masculino	38	La Paz	El Alto	PRIMAVERA B	19	Más de un año	5
Femenino	54	La Paz	La Paz	DOMINGO FAUSTINO SARMIENTO	20	Más de un año	5
Femenino	54	La Paz	La Paz	ABRAHAM REYES FE Y ALEGRIA	26	Menos de un año	0
Masculino	34	La Paz	La Paz	ABRAHAM REYES FE Y ALEGRIA	10	Más de un año	7
Masculino	35	La Paz	El Alto	TUPAC AMARU	5	Más de un año	2
Masculino	51	La Paz	La Paz	ABRAHAM REYES FE Y ALEGRIA	21	Más de un año	5
Femenino	37	La Paz	El Alto	PUERTO DE MEJILLONES CIUDAD SATELITE	13	Más de un año	3
Femenino	62	Oruro	Oruro	PANTALEON DALENCE 1	34	Más de un año	30
Femenino	44	Oruro	Oruro	JUAN MISAEAL SARACHO SECUNDARIA	19	Más de un año	11
Femenino	33	La Paz	El Alto	ABEL ITURRALDE B	5	Más de un año	3
Femenino	55	La Paz	El Alto	PRIMAVERA B	30	Más de un año	17
Masculino	38	La Paz	El Alto	1RO. DE MAYO TARDE	10	Más de un año	2
Femenino	59	Oruro	Oruro	PANTALEON DALENCE 1	25	Más de un año	21
Masculino	68	La Paz	El Alto	1RO. DE MAYO TARDE	25	Más de un año	2
Masculino	29	La Paz	La Paz	CRISTIANO VIDA NUEVA	7	Más de un año	4
Masculino	39	La Paz	El Alto	1RO. DE MAYO TARDE	8	Menos de un año	0
Masculino	55	La Paz	La Paz	ALFREDO VARGAS B	5	Menos de un año	0
Femenino	42	La Paz	El Alto	VILLA ALEMANIA	11	Más de un año	9
Masculino	30	La Paz	El Alto	BAUTISTA SAAVEDRA	11	Más de un año	6
Femenino	45	Cochabamba	Cochabamba	MERCEDES CANDIA DE OVANDO	12	Más de un año	2
Femenino	34	La Paz	El Alto	ATIPIRIS MAÑANA	8	Más de un año	5
Femenino	55	La Paz	El Alto	BAUTISTA SAAVEDRA	20	Más de un año	20
Femenino	34	La Paz	El Alto	BAUTISTA SAAVEDRA	8	Más de un año	3
Masculino	59	La Paz	El Alto	BAUTISTA SAAVEDRA	32	Más de un año	3
Masculino	47	La Paz	El Alto	BAUTISTA SAAVEDRA	22	Más de un año	13

**Fuente: Elaboración propia Excel, 2023.**

### **3.4. Aplicación del Metodología CRISP-DM.**

Para realizar correctamente el proyecto propuesto se usó el método CRISP-DM.

#### **3.4.1. Comprensión de Negocio.**

Para el estudio de clasificación del impacto de inclusión digital en las escuelas públicas, como primer punto fue comprender el proyecto como una empresa que necesita una evaluación para así verificar puntos de mejoramiento y fortalezas e identificar la problemática. Ante esto con la ayuda de la tabla 3-2, se hizo un análisis y compresión del proyecto.

**Tabla 3-2: Estrategia de Negocios.**

<b>ESTRATEGIA DE NEGOCIO</b>	
<b>Iniciativa de negocio</b>	
Identificar tipo de impacto que tuvo la inclusión digital en las escuelas públicas. Mayor información sobre los maestros y las tecnologías. Identificar requerimientos de las escuelas ante la creciente tecnología. Mejorar el proyecto	
Entidad 1	Entidad 2
Escuelas públicas de los departamentos de Bolivia.	Proyecto de inclusión Digital
<b>Decisiones clave</b>	
¿Qué departamento se encuentra con mayor impacto de inclusión digital?	¿El proyecto llegó a todo lugar?
¿Las escuelas públicas se encuentran aptas para la implementación de la inclusión digital?	¿Cómo ayudó el proyecto de inclusión digital al rendimiento en las escuelas públicas?
¿Mejorar el nivel de conocimiento tecnológico de las escuelas públicas de los departamentos de Bolivia?	¿Cómo mejorar el proyecto?
<b>Casos de uso</b>	
Predecir el tipo de impacto que causó en la educación. Ir mejorando el proyecto con nuevas tecnologías cada vez más. Lograr la participación del 100% de las escuelas públicas.	

**Fuente:** Elaboración propia, 2023.

#### **3.4.2. Comprensión de Datos.**

Realizado ya la etapa de compresión del proyecto como un negocio para identificar las problemáticas, el siguiente paso fue recolectar el conjunto de datos que fue de ayuda para alcanzar los objetivos. Para esto se usó como ya mencionado en fuentes de información secundaria el archivo llamado “Profesores\_fin” de tipo csv, que con la ayuda de Google Colab, se pudo exportar a un archivo de tipo xlsx, para una mejor comprensión de los datos, tal y como se ve en la figura 3-3.

	a0	a1	a2	a3	a4	a40	b1.b11.	b1.b12.	b1.b13.	b1.b14.	...	m2	m3	m4	m5	m6	m7
0	Sí	Masculino	32	Santa Cruz	Santa Cruz de la Sierra	NACIONAL COTOCA	Sí	No	Sí	No	...	Me faltó más por aprender.	mucho	Es muy importante para un mejor desarrollo en ...	Sí	Hubo motivación a la hora de aprender un nuevo...	excelente
1	Sí	Masculino	38	La Paz	El Alto	PRIMAVERA B	Sí	No	Sí	No	...	respondió ha algunas inquietudes que tuve como...	mucho	al tener conocimiento suficiente de las herramientas...	Sí	aun estamos en proceso de manejar estos softwa...	bueno
2	Sí	Femenino	54	La Paz	La Paz	DOMINGO FAUSTINO SARMIENTO	Sí	No	Sí	Sí	...	Por que nos permite generar material para el desarrollo...	mucho	Porque rompe con la Monotonía de una clase tradicional...	Sí	Nos permite acceso ilimitado y económico.	excelente
3	Sí	Femenino	54	La Paz	La Paz	ABRAHAM REYES FE Y ALEGRIA	No	No	Sí	No	...	Porque nos enseñaron con interés y paciencia.	mucho	Porque tenemos que motivar a los estudiantes...	Sí	Porque se motivaron más.	bueno
4	Sí	Masculino	34	La Paz	La Paz	ABRAHAM REYES FE Y ALEGRIA	No	No	Sí	No	...	Siempre estuvieron atentos a nuestras dudas y...	mucho	Se diversifica las estrategias didácticas y...	Sí	Motiva al estudiante a conocer y profundizar más...	excelente

**Figura 3-3: Conjunto de datos “Profesores\_fin” en formato csv.**

**Fuente:** Elaboración propia Google Colab, 2023.

Tal como se muestra en la figura 3-3, el conjunto de datos se encuentra con nombres de variables como a0, a1, a2, etc, lo cual no da a entender bien a qué hace referencia; para esto con la ayuda de Google Colab se puede exportarlo en formato xlsx (Excel), para así poder manipularlo de mejor manera y realizar el cambio de los nombres de las variables por otros que sean más entendibles como sexo, edad, ciudad, departamento, etc, tal y como se ve en la tabla 3-3. Se eliminaron columnas que no aportaban mucha información al estudio llegando a reducirse a solo 103.

**Tabla 3-3: Conjunto de datos “Profesores\_final” en formato xlsx.**

Sexo	Edad	Departamento	Ciudad	Colegio	Años de experiencia
Masculino	32	Santa Cruz	Santa Cruz de la Sierra	NACIONAL COTOCA	20
Masculino	38	La Paz	El Alto	PRIMAVERA B	19
Femenino	54	La Paz	La Paz	DOMINGO FAUSTINO SARMIENTO	20
Femenino	54	La Paz	La Paz	ABRAHAM REYES FE Y ALEGRIA	26
Femenino	37	La Paz	El Alto	PUERTO DE MEJILLONES CIUDAD SATELITE	13
Femenino	62	Oruro	Oruro	PANTALEON DALENCE 1	34
Femenino	44	Oruro	Oruro	JUAN MISael SARACHO SECUNDARIA	19
Femenino	33	La Paz	El Alto	ABEL ITURRALDE B	5

**Fuente:** Elaboración propia, 2023.

El conjunto de datos muestra muchos aspectos y características de los profesores y las escuelas donde imparten clases, pero se vio óptimo ordenarlos para comprenderlo de mejor forma.

### **3.4.2.1. Datos personales de los profesores.**

La tabla 3-4, muestra las variables y los valores que toman de los datos personales de los profesores que participaron en la encuesta.

**Tabla 3-4: Tabla de datos personales de los profesores.**

Descripción	Variables	Valores
Datos personales de los profesores	Sexo	Hombre, mujer
	Edad	20-78
	Ciudad	16 ciudades
	Departamento	9 departamentos
	Colegio	Varios
	Años de experiencia	1-45

Fuente: Elaboración propia Word, 2023.

### **3.4.2.2. Datos de nivel académico de los profesores.**

La tabla 3-5, muestra el nivel académico de los profesores que participaron de la encuesta del proyecto inclusión digital.

**Tabla 3-5: Tabla de datos de nivel académico de los profesores.**

Descripción	Variables	Valores
Nivel académico de los profesores.	Normalista	Si, No
	Técnico	Si, No
	Licenciatura	Si, No
	Especialidad	Si, No
	Maestría	Si, No
	Doctorado	Si, No

Fuente: Elaboración propia Word, 2023.

### **3.4.2.3. Datos de materias que imparten los profesores.**

En la tabla 3-6, se puede observar las materias que dictan los profesores.

**Tabla 3-6: Tabla de datos de materias que imparten los profesores.**

Descripción	Variables	Valores
Materias que imparten los profesores	Matemáticas	Si, No
	Técnica Tecnológica	Si, No
	Ciencias Naturales	Si, No
	Geografía	Si, No
	Biología	Si, No
	Física	Si, No
	Química	Si, No
	Lenguaje	Si, No
	Lengua extranjera	Si, No
	Ciencias Sociales	Si, No
	Artes plásticas	Si, No
	Música	Si, No
	Religión	Si, No
	Filosofía	Si, No

Fuente: Elaboración propia Word, 2023.

#### **3.4.2.4. Datos de herramientas de apoyo de los profesores.**

En la tabla 3-7, se encuentran las herramientas que sirven de apoyo a los profesores tanto para la educación e impartir sus clases con mayor provecho.

**Tabla 3-7: Tabla de herramientas de apoyo de educación de los profesores.**

Descripción	Variables	Valores
Herramientas de apoyo de los profesores	Computadora	Si, No
	Cantidad de computadoras	0-7
	Tablets	Si, No
	Cantidad de Tablets	0-4
	Celular	Si, No
	Celular inteligente	Si, No

Fuente: Elaboración propia Word, 2023.

#### **3.4.2.5. Datos de utilidad de computadora y celular.**

En la tabla 3-8, se muestra la utilidad que tiene tanto una computadora y un celular inteligente, para un profesor.

**Tabla 3-8: Tabla que muestra la utilidad de las herramientas como computadora y celular.**

Descripción	Variables	Valores
Utilidad de computadora y el celular	Planillas	Si, No
	Juegos	Si, No
	Estudio	Si, No
	Música	Si, No
	Redes Sociales	Si, No
	Planificar clases	Si, No
	Uso en clases	Si, No
	Videos	Si, No
	Internet	Si, No
	Llamar	Si, No
	Fotos	Si, No

Fuente: Elaboración propia Word, 2023.

#### **3.4.2.6. Datos de acceso y utilidad del internet en los profesores.**

En la tabla 3-9, se observa el acceso y la utilidad de este servicio vital para la educación de hoy en día; el contar con acceso a internet en todo lugar mantiene a la población conectada sin importar en qué lugar se encuentren y hace posible acceder a toda la información que internet puede ofrecer a la educación de las escuelas.

**Tabla 3-9: Tabla de acceso y utilidad del internet en los profesores.**

Descripción	Variables	Valores
Acceso y utilidad del internet en los profesores.	Internet en casa	Si, No
	Internet en celular	Si, No
	Internet en colegio	Si, No
	Fines académicos	Si, No
	Descargas	Si, No
	Redes Sociales	Si, No
	Búsqueda de información	Si, No
	Trabajo	Si, No
	Jugar	Si, No
	Música	Si, No

Fuente: Elaboración propia Word, 2023.

#### **3.4.2.7. Datos de uso de Redes Sociales.**

En la tabla 3-10, se muestra que redes sociales son utilizadas por los maestros como apoyo y colaborador para sus métodos de enseñanza.

**Tabla 3-10: Tabla de uso de redes sociales de los profesores.**

Descripción	Variables	Valores
Uso de redes sociales	Facebook	Si, No
	YouTube	Si, No
	WhatsApp	Si, No
	Messenger	Si, No
	Instagram	Si, No
	Telegram	Si, No
	TikTok	Si, No

Fuente: Elaboración propia Word, 2023.

#### **3.4.2.8. Datos de conocimientos de los profesores.**

En la tabla 3-11, se muestran los conocimientos de herramientas de Google que tienen los profesores, también los conocimientos básicos en informática y ofimática denotados con saber 1, 2 y 3; y como un detalle que hace hoy en día a los maestros más eficientes es conocer lo que son las TIC, y también si logran implementar en sus métodos de enseñanza.

**Tabla 3-11: Conocimientos e implementaciones de los profesores para la educación.**

Descripción	Variables	Valores
Conocimientos de los profesores	Google Académico	Si, No
	Google Buscador	Si, No
	Google Drive	Si, No
	Correo electrónico	Si, No
	Saber 1	Si, No, Dudososo
	Saber 3	Si, No, Dudososo
	Saber 5	Si, No, Dudososo
	Conocimientos en TIC	Si, No
	Incorporación de TIC	Casi siempre, Nunca, Ocasionalmente, rara vez, siempre
	Disponibilidad de enseñanza TIC	Si, No

Fuente: Elaboración propia Word, 2023.

### **3.4.2.9. Datos de formas de preparación de maestros y estudiantes para las clases.**

La tabla 3-12, muestra de qué manera o que método utilizan los maestros para elaborar su material de enseñanza, y que métodos utilizan los estudiantes para estudiar.

**Tabla 3-12: Tabla de formas de preparación de los maestros y estudiantes.**

Descripción	Variables	Valores
Formas de preparación de maestros.	Elaboración de material de internet	Si, No
	Elaboración propia	Si, No
	Elaboración por trabajo compartido	Si, No
	Modalidad de clases	Clases frontales, Clases prácticas, laboratorios, seminarios, talleres
Formas de preparación de los estudiantes.	Por videos	Si, No
	Por presentaciones	Si, No
	Por textos en línea	Si, No
	Por textos impresos	Si, No

Fuente: Elaboración propia Word, 2023.

### **3.4.2.10. Datos de requerimientos de las escuelas.**

En la tabla 3-13, se muestra los requerimientos que las escuelas necesitan para así poder mejorar la calidad de educación e infraestructura.

**Tabla 3-13: Tabla de requerimientos de las escuelas.**

Descripción	Variables	Valores
Requerimientos de las escuelas.	Mobiliario	Si, No
	Medidas de seguridad	Si, No
	Equipo computacional	Si, No
	Internet	Si, No

Fuente: Elaboración propia Word, 2023.

### **3.4.2.11. Datos de características de los colegios.**

En la tabla 3-14, se muestra algunas de las características que algunas escuelas tienen.

**Tabla 3-14: Tabla de características de los colegios.**

Descripción	Variables	Valores
Característica de las escuelas.	Uso de celular en clases	Si, No
	Aula computacional	Si, No
	Acceso a internet en alguna aula	Si, No

Fuente: Elaboración propia Word, 2023.

### **3.4.3. Preparación de los Datos.**

La preparación de datos es una etapa donde ya se comienza a entender el comportamiento de estos datos y que información ofrece el conjunto de datos analizados. Para esto se realizó un Análisis Exploratorio

de Datos (AED), y con la ayuda de Google Colaboratory se realizó el análisis con códigos de programación en lenguaje Python.

### **3.4.3.1. Configuración Inicial.**

Con la ayuda de importación de librerías de Python y cargando el conjunto de datos (dataset) desde Drive, se crea un DataFrame (Marco de Datos) donde se almacena todo el conjunto de datos, como se ve en la figura 3-4.

```
#Importa las librerias necesarias
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

from google.colab import drive
drive.mount('/content/drive')

# Acceder al archivo en tu unidad de Google Drive
ruta_archivo = '/content/drive/MyDrive/profesores_final.xlsx'

#creacion del DataFrame
df = pd.read_excel(ruta_archivo)
```

**Figura 3-4: Configuración Inicial en Google Colab.**

Fuente: Elaboración propia Google Colab, 2023.

### **3.4.3.2. Exploración Inicial de los Datos.**

En la figura 3-5, 3-6, 3-7, 3-8, se puede observar ya los primeros registros y algunas variables del conjunto de datos mediante los comandos que se muestran en la tabla 3-15.

**Tabla 3-15: Comandos de Python para exploración inicial de Datos.**

Comandos	Descripción
df.head()	Permite visualizar las primeras filas de tu DataFrame
df.info()	Proporciona información más detallada, incluyendo tipos de datos y la presencia de valores nulos
df.shape	Otorga las dimensiones del DataFrame en términos de cantidad de filas y columnas
df.dtypes	Muestran los tipos de datos de cada columna en un DataFrame

Fuente: Elaboración propia, 2023.

```
#Muestra los primeros registros del DataFrame  
df.head()
```

	Sexo	Edad	Departamento	Ciudad	Colegio	Años de experiencia	Tiempo en el colegio	Años en el Colegio	Normalista	Tecnico
0	Masculino	32	Santa Cruz	Santa Cruz de la Sierra	NACIONAL COTOCOCA	20	Más de un año	15	Sí	No
1	Masculino	38	La Paz	El Alto	PRIMAVERA B	19	Más de un año	5	Sí	No
2	Femenino	54	La Paz	La Paz	DOMINGO FAUSTINO SARMIENTO	20	Más de un año	5	Sí	No

**Figura 3-5: Comando head() para visualizar el DataFrame.**

Fuente: Elaboración propia Google Colab, 2023.

```
#Obtencion de informacion sobre el DataFrame  
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 871 entries, 0 to 870  
Columns: 103 entries, Sexo to Uso de tecnologias mejoro el aprendizaje  
dtypes: int64(6), object(97)  
memory usage: 701.0+ KB
```

**Figura 3-6: Comando info() para ver las cantidades .**

Fuente: Elaboración propia Google Colab, 2023.

```
#Obtencion del tamaño del DataFrame  
df.shape
```

```
(871, 103)
```

**Figura 3-7: Comando shape.**

Fuente: Elaboración propia Google Colab, 2023.

```
df.dtypes
```

```
Sexo                      object  
Edad                     int64  
Departamento              object  
Ciudad                    object  
Colegio                   object  
                           ...  
Requerimiento de Mobiliario    object  
Requerimiento de equipo computacional    object  
Requerimiento de internet       object  
Calificacion del proyecto     object  
Uso de tecnologias mejoro el aprendizaje    object  
Length: 103, dtype: object
```

**Figura 3-8: Comando dtypes.**

Fuente: Elaboración propia Google Colab, 2023.

Como se puede observar en la figura 3-8, cuenta con bastantes variables de tipo object que hacen referencia a valores de escritura y las variables de tipo int64 que hace referencia a valores numéricos. También en la figura 3-7, muestra que el tamaño es de 871 registros (filas) y 103 columnas.

### 3.4.3.3. Estadísticas descriptivas

En la figura 3-9, tomando en cuenta solo valores numéricos presentes en el conjunto de datos, se observa valores importantes como media, mediana, moda, cuartiles, desviación estándar, límites máximo y mínimo, etc, los cuales nos ayudan a comprender mejor los datos.

```
#Para obtener estadistica descriptiva basicas
df.describe()
```

	Edad	Años de experiencia	Años en el Colegio	Cantidad de computadoras	Cantidad de Tablets	Días en la semana usando internet
<b>count</b>	871.000000	871.000000	871.000000	871.000000	871.000000	871.000000
<b>mean</b>	44.463835	17.462687	8.252583	1.416762	0.314581	5.146958
<b>std</b>	9.755200	9.356695	7.327513	0.760909	0.579144	2.471723
<b>min</b>	20.000000	1.000000	0.000000	0.000000	0.000000	0.000000
<b>25%</b>	37.000000	10.000000	3.000000	1.000000	0.000000	4.000000
<b>50%</b>	44.000000	16.000000	6.000000	1.000000	0.000000	7.000000
<b>75%</b>	51.000000	24.000000	12.000000	2.000000	1.000000	7.000000
<b>max</b>	78.000000	45.000000	44.000000	7.000000	4.000000	7.000000

**Figura 3-9: Estadística descriptiva de los datos.**

Fuente: Elaboración propia Google Colab, 2023.

### 3.4.3.4. Visualizaciones.

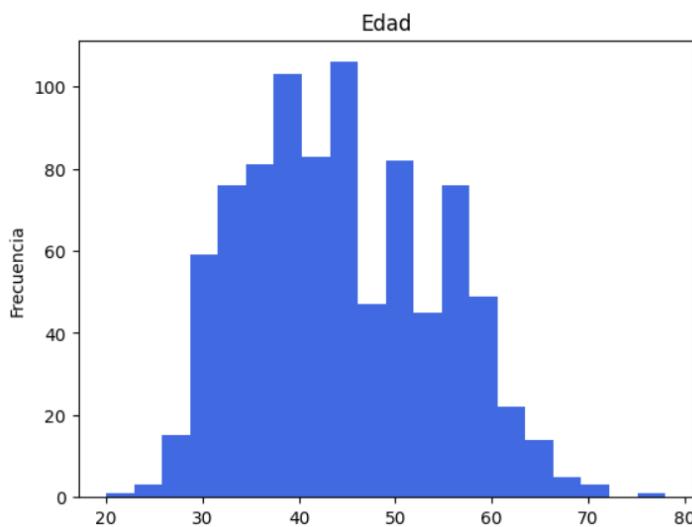
Usando graficas como histogramas y diagramas de barras se puede observar el comportamiento de las variables de nuestro conjunto de datos, con el cual se observó visualmente a aquellos puntos que muy fácilmente no se pueden apreciar.

En la figura 3-10 y 3-11 se muestra la gráfica de histograma de la variable edad, donde se distingue la variabilidad de edades de los maestros que fueron participes de la encuesta del programa inclusión digital. Claramente nos muestra como las edades presentes de los maestros en su mayoría son aproximadamente entre 30 y 60 años.

```
#Grafico de Histograma para la variable Edad
plt.hist(df['Edad'], bins=20, color="#4169E1")
plt.ylabel('Frecuencia')
plt.title('Edad')
plt.show()
```

**Figura 3-10: Código Python para el Histograma de Edad.**

Fuente: Elaboración propia Google Colab, 2023.

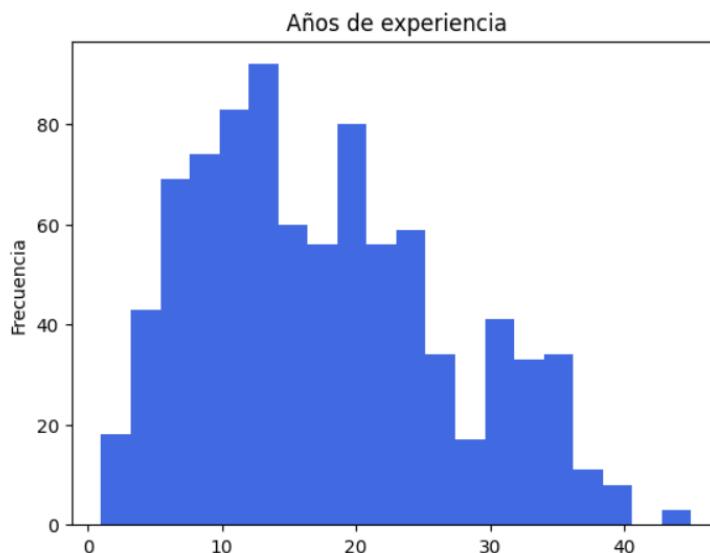


**Figura 3-11: Histograma de Edad.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-12, se muestra la gráfica del histograma de la variable años de experiencia, donde es notorio diversidad de años de experiencia. Nos da a entender que los años de experiencia con mayor frecuencia de los maestros es de aproximadamente entre 5 a 25 años.

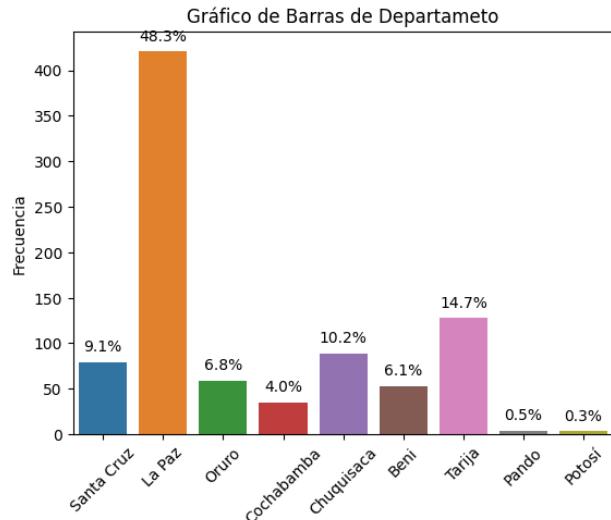
```
#Grafico de histograma para la variable Años de experiencia
plt.hist(df['Años de experiencia'], bins=20, color="#4169E1")
plt.ylabel('Frecuencia')
plt.title('Años de experiencia')
plt.show()
```



**Figura 3-12: Histograma de Años de experiencia.**

Fuente: Elaboración propia Google Colab, 2023.

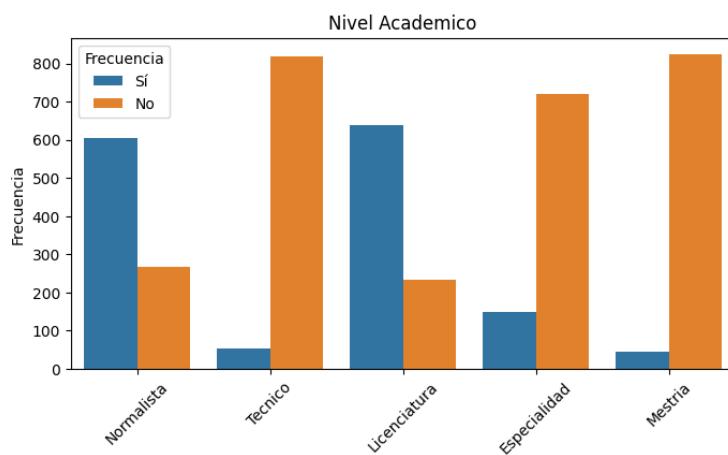
En la figura 3-13, se muestra mediante diagrama de barras la variable departamento, donde visiblemente se ve que el departamento de La Paz, tuvo mayor participación en la encuesta del programa inclusión digital con 48,3%, seguido por el departamento de Tarija con 14,7% de participación, seguidos por los departamentos de Chuquisaca, Santa Cruz, Oruro, Beni, Cochabamba, Pando y Potosí.



**Figura 3-13: Diagrama de barras de la variable Departamento.**

**Fuente:** Elaboración propia Google Colab, 2023.

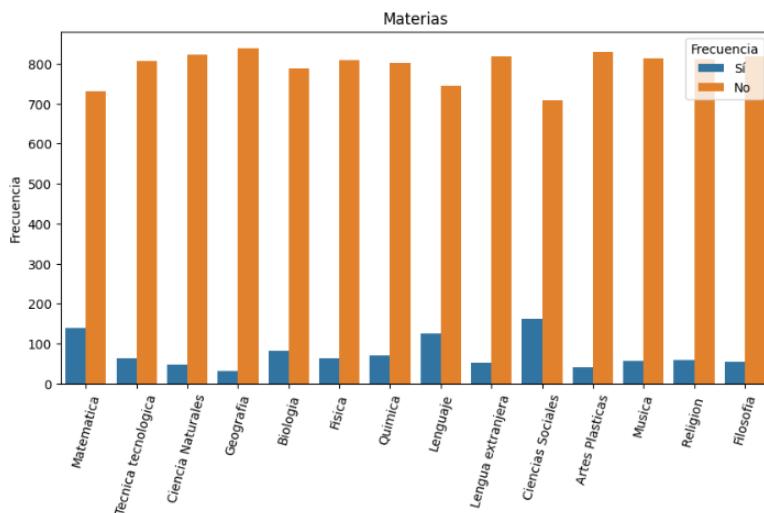
En la figura 3-14, se muestra por medio de diagrama de barras la frecuencia del nivel académico presente de los maestros que fueron partícipes del programa inclusión digital. Donde es notorio que los maestros en su mayoría son educación en licenciatura, seguidos por normalista, que hoy en día es el nivel de educación que todos los maestros obtiene al terminar su educación, y son pocos que continúan estudiando y alcanzar niveles más altos de educación como especialidad, maestría y doctorado.



**Figura 3-14: Diagrama de barras de Nivel Académico.**

**Fuente:** Elaboración propia Google Colab, 2023.

En la figura 3-15, mediante diagrama de barras de la variable Materia, se observa que en mayor frecuencia los maestros participes fueron de todas las asignaturas pero en poca cantidad, y las asignaturas mas sobresalientes en cantidad fueron Ciencias Sociales, Lenguaje, Matemáticas, Biología, que llegan a alcanzar entre 120 a 170 maestros por materia.

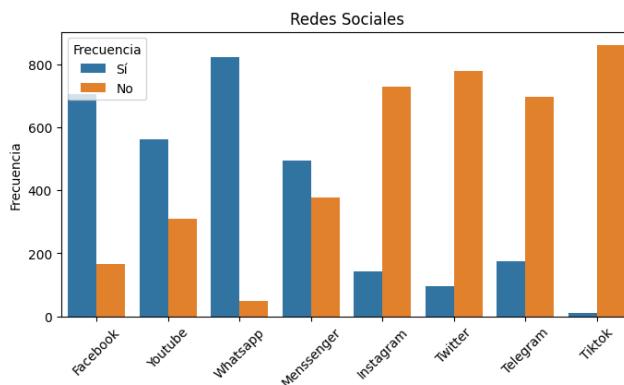


**Figura 3-15: Diagrama de barras de Materias dictadas por los maestros.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-16, por medio de diagrama de barras de la variable redes sociales, se observa que los maestros participes en la encuesta son conocedores y usan las redes sociales, donde la red social que más destaca entre todas es WhatsApp, llegando casi al 100% en uso de todos los maestros, seguidos por la red social como Facebook, YouTube, Messenger, Telegram, Instagram, Twitter y Tiktok.

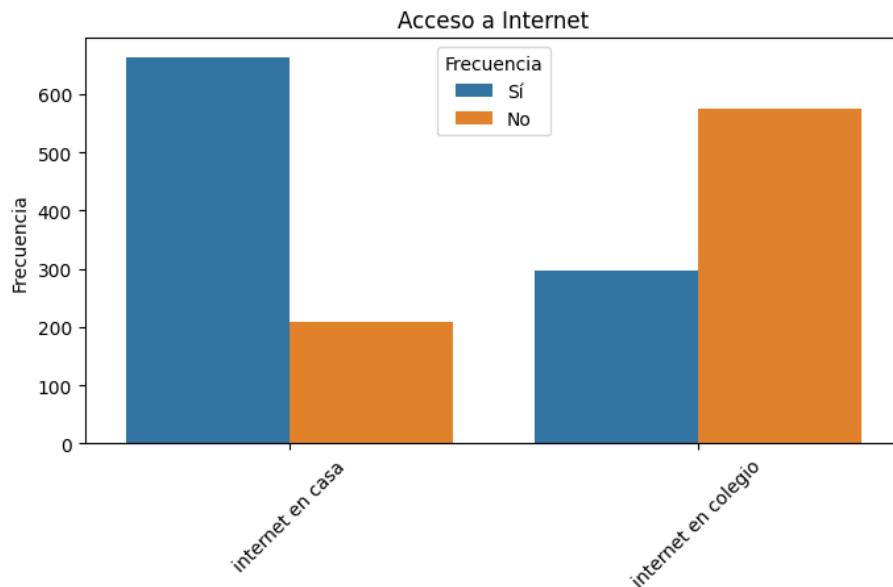
Las redes sociales hoy en día son consideradas también herramientas de enseñanza, ya que no es necesario la presencia del maestro ni del estudiante en un mismo lugar, sino que se lo puede hacer todo remotamente, en cualquier lugar que se disponga de conexión a internet.



**Figura 3-16: Diagrama de barras de Redes Sociales.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-17, por medio del diagrama de barras de la variable Acceso a internet, claramente se ve que los maestros participes de la encuesta tienen acceso a internet en mayor frecuencia en sus hogares, más en sus escuelas son pocos maestros que cuentan con este servicio tan esencial que sirve como apoyo y beneficio de la educación para sus estudiantes.



**Figura 3-17: Diagrama de barras de Acceso a internet.**

Fuente: Elaboración propia Google Colab, 2023.

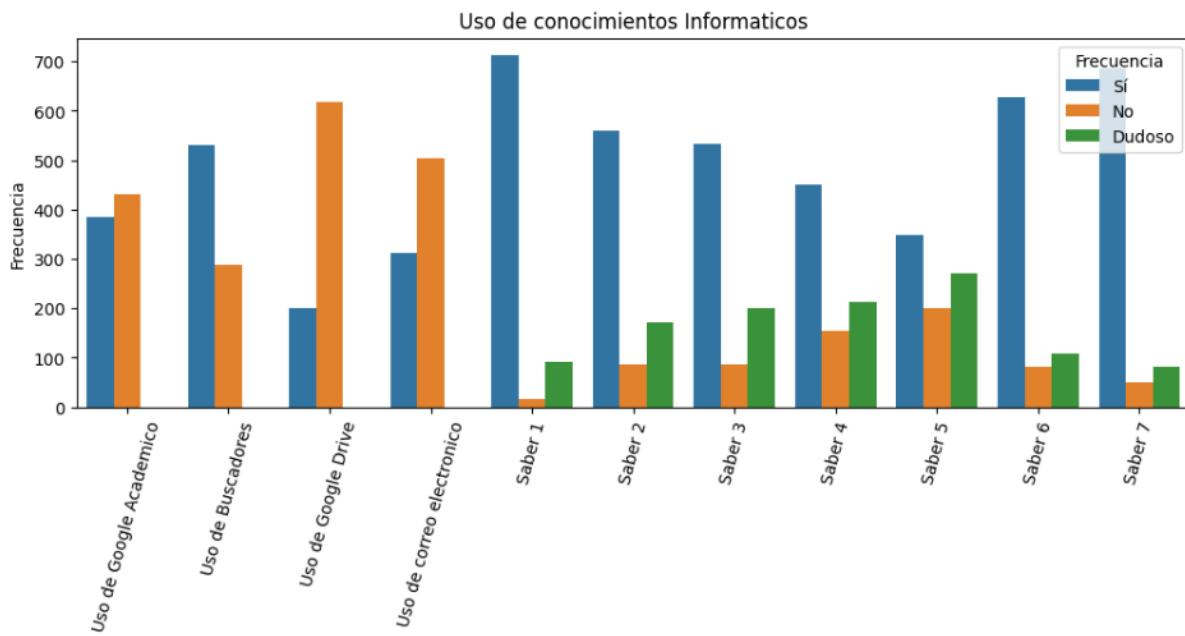
Un punto muy importante hoy en día es la conexión a internet, más que una necesidad se podría decir que llegaría a ser un derecho el tener internet en las unidades educativas, ya que, sin este servicio, los conocimientos las herramientas y demás quedan sin sentido porque todo es dependiente de este.

En la figura 3-16, se observa que en gran parte los maestros cuentan con internet en sus hogares, pero donde debería estar al 100% y activo es en las unidades educativas, cosa que no es así según las gráficas resultantes.

En la figura 3-17, con el uso de diagramas de barras de la variable conocimientos informáticos, se observa los conocimientos de los maestros con respecto a herramientas de Google y conocimientos informáticos.

Es notorio que los maestros son poco conocedores y regularmente bien las herramientas de Google como ser Google Académico, Google Drive, correo electrónico, y solo destaca con mayor conocimiento el uso de buscador de Google.

En cuestión de conocimientos básicos de informática los maestros destacan y se presentan con mayor frecuencia que si los maestros saben usar los conocimientos que se ven en la tabla 3-4.



**Figura 3-18: Diagrama de barras de usos de conocimientos Informáticos.**

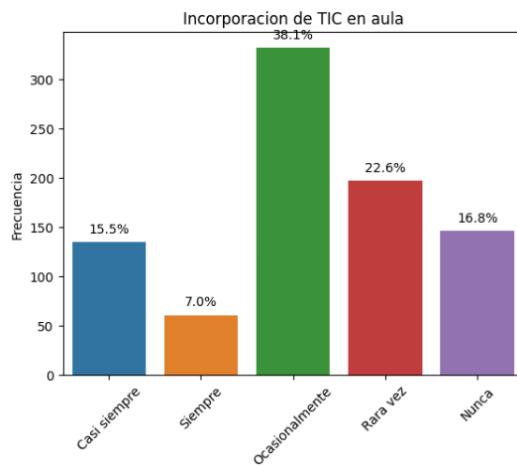
Fuente: Elaboración propia Google Colab, 2023.

**Tabla 3-16: Descripción del significado de las variables saber 1,2,3,4,5,6,7.**

Saberes	Descripción
Saber 1	elaborar, copiar o desplazar texto en un documento y/o buscar, copiar y mover archivos
Saber 2	enviar y abrir correos electrónicos con archivos adjuntos?
Saber 3	conectar o instalar nuevos dispositivos y/o transferir archivos entre la computadora y otro dispositivo
Saber 4	buscar, instalar y configurar programas y/o aplicaciones desde su celular?
Saber 5	buscar, instalar y configurar programas y/o aplicaciones desde su computadora?
Saber 6	conectarse a una red wifi desde una laptop
Saber 7	conectarse a una red wifi desde un teléfono móvil

Fuente: Elaboración propia, 2023.

En la figura 3-19, mediante diagrama de barra de la variable incorporacion de TIC en aula, claramente se observa como los maestros de las escuelas publicas incorporan de forma ocasional y en un porcentaje de 38,1% las TIC en sus aulas, y de forma Siempre solo en un 7%, dando a conocer con esta información que esta sucediendo algo o hay algun impedimento que hace que los maestros no implementen las TIC de forma cotidiana todos los dias o en todas sus materias.



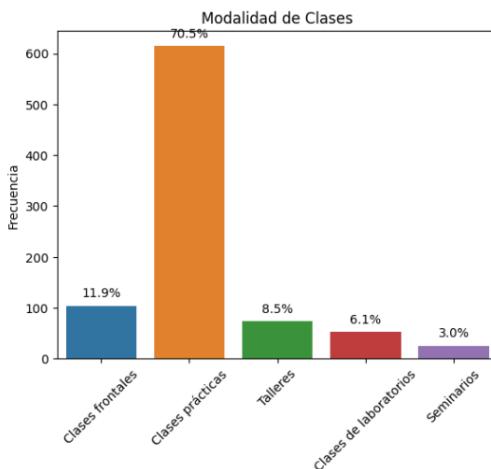
**Figura 3-19: Diagrama de barras de Incorporación de TIC.**

Fuente: Elaboración propia Google Colab, 2023.

El uso de TIC en las unidades educativas se ha vuelto algo necesario y escencial, todos los maestros sin importar su edad deberían poder ser capaces de adecuar las TIC a su enseñanza, para esto el gobierno de Bolivia a estado lanzando capacitaciones regularmente y más con el proyecto de inclusión digital.

La grafica da información de que las TIC todavía no están siendo implementadas al 100% sino solamente ocasionalmente, rara vez o nunca, esto es algo preocupante en la actualidad.

En la figura 3-20, por medio del diagrama de barras de la variable modalidad de clases, evidentemente se observa como la modalidad de clases que los maestros hoy en día utilizan es de forma práctica, rompiendo así la enseñanza clásica de hace mucho tiempo atrás. El 70.5% nos indica que los maestros evaluados en la encuesta afirmaron que realizan sus clases de forma práctica, seguida por clases frontales, talleres, laboratorios y seminarios.

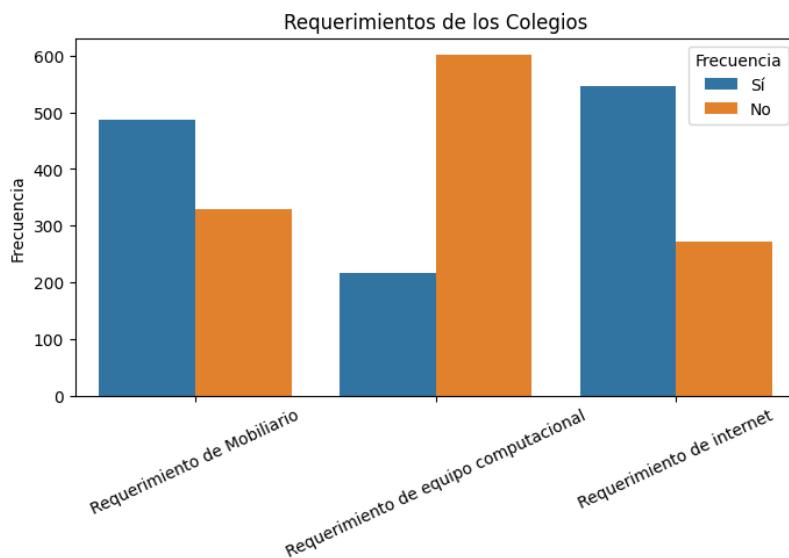


**Figura 3-20: Diagrama de Modalidades de Clases.**

Fuente: Elaboración propia Google Colab, 2023.

Pese que las gráficas anteriores dan indicio que la educación seguía de forma clásica, esta grafica contrasta y da a entender que más del 70% de los maestros imparten sus clases de forma práctica.

En la figura 3-21, por medio del diagrama de barras de la variable requerimiento de los colegios, se puede verificar que en su mayoría los colegios solicitan tener acceso al servicio de internet, ya que carecen de comunicación con la tecnología, y tambien seguido por requerimiento de mobiliario y por ultimo equipos computacionales.



**Figura 3-21: Diagrama de barras de Requerimientos de los Colegios.**

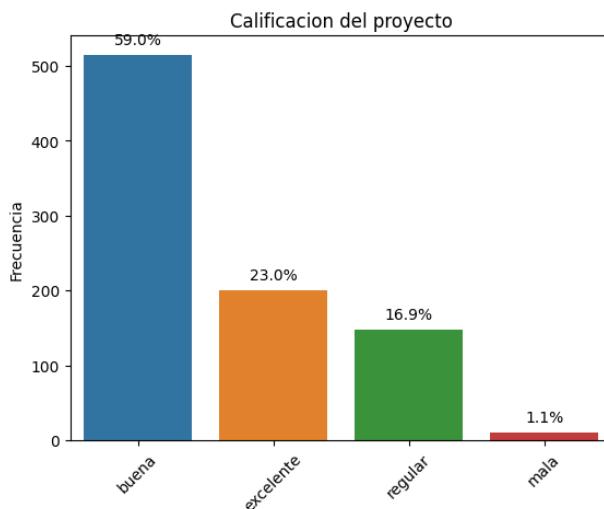
Fuente: Elaboración propia Google Colab,2023.

Es muy notable como los colegios tienen la necesidad de poder contar con el servicio de acceso a internet, ya que sin este es en vano todo el aprendizaje de herramientas de apoyo para la educación como ser TIC, redes sociales y demás.

El requerimiento de muebles también es algo importante en los colegios y no dejando de lado el poder contar con equipos computacionales como ser una computadora para cada estudiante.

En la figura 3-22, por medio de diagrama de barras de la variable calificación del proyecto de inclusión digital, podemos ver que los maestros de todos los departamentos de Bolivia dieron como calificación al proyecto como Bueno en un 59%. Se podria decir que el programa esta en un rumbo bueno pero la

calificación de excelente como solo un 23%, da a pensar mucho el por que solo algunos maestros lograron dar la calificación como excelente. Se podria pensar que algo necesita mejorar en el programa.



**Figura 3-22: Diagrama de barras de Requerimientos de los Colegios.**

Fuente: Elaboración propia Google Colab, 2023.

Los maestros calificaron al proyecto de inclusión digital como simplemente bueno, pero hoy en día decir bueno es como decir que faltó algo más, dando un indicio que el proyecto necesita mejorar cada vez más y así poder alcanzar una calificación Excelente, es muy posible que se debe mejorar en cuestión de capacitores, material de capacitación, horarios, tiempo, etc. Para ver todas las gráficas ir a Anexos 2 y 3

#### 3.4.3.5. Manejo de Datos Faltantes.

El poder llegar a verificar y corregir valores faltantes o valores nulos dentro del conjunto y datos de los maestros y sus escuelas, nos ayudó a poder evitar errores futuros para nuestro modelo de machine learning.

Como en la figura 3-23, por medio del comando `df.isnull.sum()`, podemos ver que variables cuentan con valores nulos, pero para nuestro caso no es apreciable todas las variables con valores nulos.

```
df.isnull().sum()

Sexo          0
Edad          0
Departamento 0
Ciudad        0
Colegio       0
...
Requerimiento de Mobiliario    0
Requerimiento de equipo computacional 0
Requerimiento de internet        0
Calificación del proyecto      0
Uso de tecnologías mejora el aprendizaje 0
Length: 103, dtype: int64
```

**Figura 3-23: Verificación de valores nulos.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-24, por medio del mapa de calor, se observa como gráficamente podemos observar valores nulos de nuestras variables, donde los valores nulos están pintados de color naranja, y por este medio se pudo verificar de forma más sencilla y exacta que existen valores nulos.



**Figura 3-24: Mapa de Calor.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-25, por medio de comandos Python en Google Colab podemos ver el nombre y la cantidad de los valores nulos existentes en nuestro conjunto de datos de los maestros y sus escuelas.

```
# Obtener la cantidad de valores nulos por columna
nulos_por_columna = df.isnull().sum()

# Imprimir el nombre de la columna y la cantidad de valores nulos
for columna, cantidad_nulos in nulos_por_columna.items():
    print(f"{columna}: {cantidad_nulos} valores nulos")

Tipo de computadora: 18 valores nulos
tipo de internet: 115 valores nulos
lugar de acceso a internet: 115 valores nulos
```

**Figura 3-25: Valores Nulos.**

Fuente: Elaboración propia Google Colab, 2023.

Una vez descubierto los valores nulos, se procedió a tratarlos para no tener que eliminarlos, ya que se perdería información muy valiosa.

En la figura 3-26, por medio de comandos Python en Google Colab, se ve como los valores nulos de las variables Tipo de computadora, tipo de internet y lugar de acceso a internet, son remplazados por valores iguala Ninguna.

```
df['Tipo de computadora'] = df['Tipo de computadora'].fillna('Ninguna')

df['tipo de internet'] = df['tipo de internet'].fillna('Ninguna')

df['lugar de acceso a internet'] = df['lugar de acceso a internet'].fillna('Ninguno')
```

**Figura 3-26: Tratamiento de Valores Nulos.**

Fuente: Elaboración propia Google Colab, 2023.

#### 3.4.3.6. Manejo de Valores Atípicos.

En la figura 3-27, por medio de comandos como df.shape vemos los valores de los registros igual a 871, pero después de aplicar la fórmula de Z-score para eliminar valores atípicos solo se cuenta con valores de registros igual a 818, dando a entender que si existían 53 valores atípicos.

```
df.shape
(871, 103)

# Definir un umbral para identificar valores atípicos (por ejemplo, z-score > 3 o < -3)
umbral = 3

# Iterar a través de las columnas numéricas
for columna in df.select_dtypes(include=[np.number]):
    # Calcular z-scores para la columna actual
    z_scores = (df[columna] - df[columna].mean()) / df[columna].std()

    # Filtrar los datos originales para eliminar los valores atípicos en la columna actual
    df = df[(z_scores < umbral) & (z_scores > -umbral)]
```

```
df.shape
(818, 103)
```

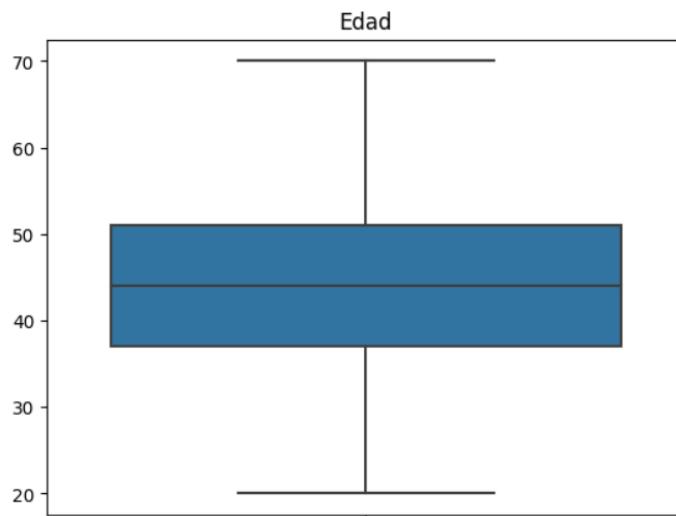
**Figura 3-27: Eliminación de Valores Atípicos.**

Fuente: Elaboración propia Google Colab, 2023.

#### 3.4.3.7. Análisis Univariado y Bivariado.

Este análisis es de mucha utilidad para entender el comportamiento de 2 o más variables entre sí.

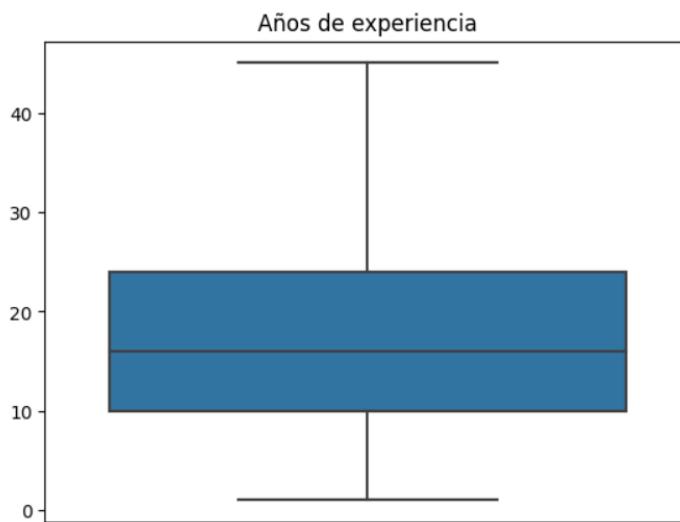
En la figura 3-28, por medio de diagramas de cajas de la variable edad, se observa con más claridad que los límites de edades en forma superior es de 70 años y como límite inferior es de 20 años, siendo el rango de edades con mayor frecuencia entre 38 a 51 años.



**Figura 3-28: Diagrama de cajas de Edad.**

Fuente: Elaboración propia Google Colab, 2023.

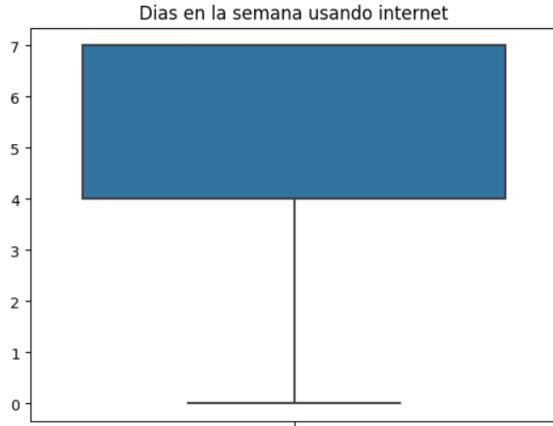
En la figura 3-29, de igual forma mediante diagramas de cajas por código Python en Google Colab, se puede observar los límites de años de experiencia de los maestros que van desde 0 años hasta 45 años aproximadamente, y los años de experiencia con mayor frecuencia se encuentra entre 10 a 25 años.



**Figura 3-29: Diagrama de cajas de Años de experiencia.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-30, mediante el diagrama de cajas de la variable “Días en la semana usando internet”, se puede observar claramente que los maestros usan internet entre 4 a 7 días de la semana, siendo 7 el valor más frecuente que dijeron que los maestros usan internet.

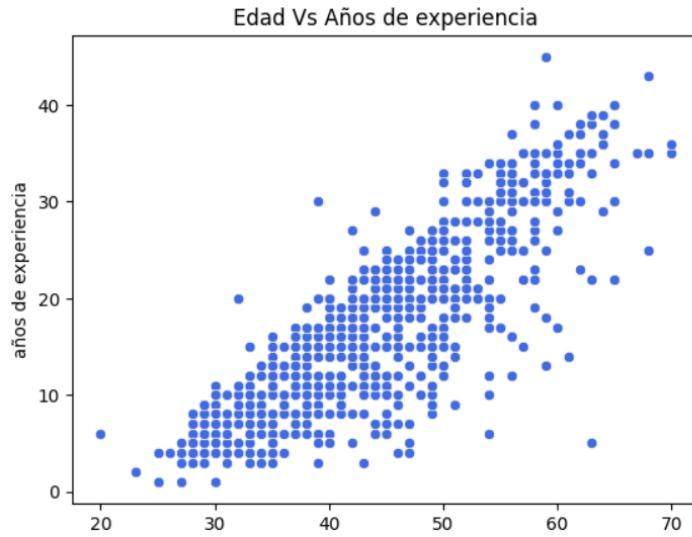


**Figura 3-30: Diagrama de días en la semana usando internet.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-31, mediante el comando scatterplot podemos observar la relación entre las variables edad y años de experiencia, dando como resultado relación lineal positiva entre las dos variables, indicando que si existe relación entre estas dos variables.

```
#diagrama de dispersión entre 'edad' y 'años de experiencia'
sns.scatterplot(data=df, x='Edad', y='Años de experiencia', color="#4169E1")
plt.xlabel('edad')
plt.ylabel('años de experiencia')
plt.title('Edad Vs Años de experiencia')
plt.show()
```



**Figura 3-31: Diagramas de correlación Edad Vs Años de Experiencia.**

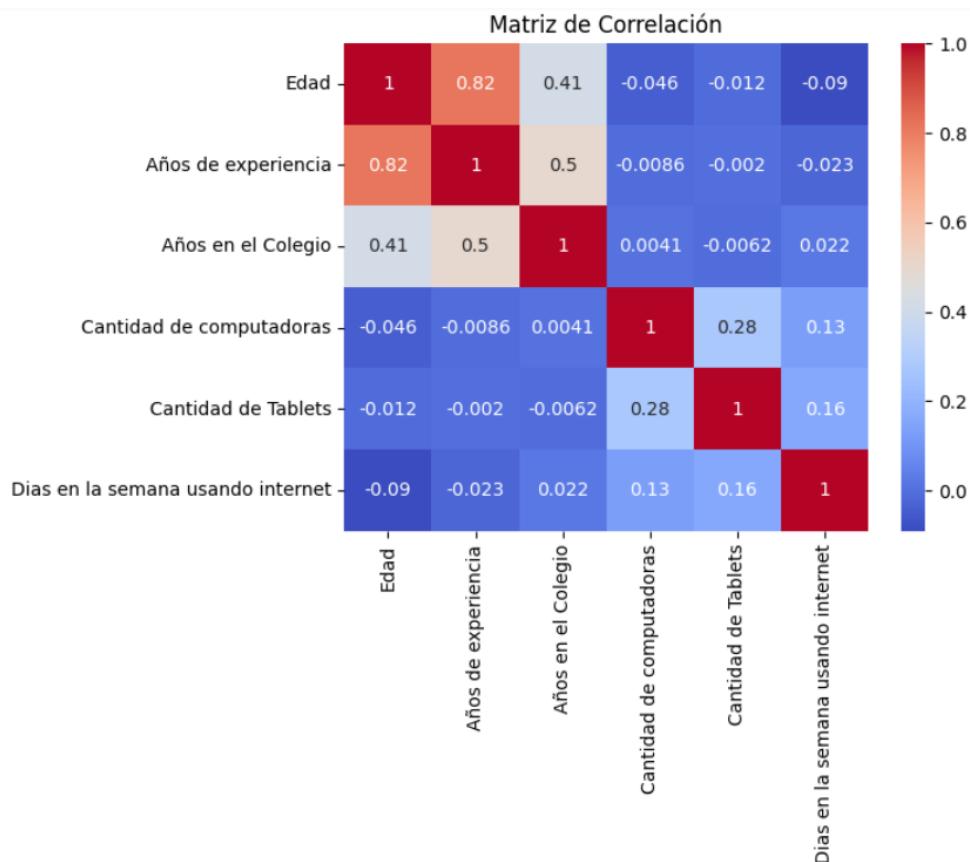
Fuente: Elaboración propia Google Colab, 2023.

Para una mejor revisión de la relación entre las variables edad y años de experiencia la figura 3-32 y 3-33, nos muestra mediante los comando df.corr() y sns.heatman una matriz de correlación.

```
matriz_correlacion = df.corr()  
sns.heatmap(matriz_correlacion, annot=True, cmap='coolwarm')  
plt.title('Matriz de Correlación')  
plt.show()
```

**Figura 3-32: Código de matriz de correlación.**

Fuente: Elaboración propia Google Colab, 2023.



**Figura 3-33: Matriz de Correlación.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-33 se muestra que la relación entre edad y años de experiencia es de valor 0,82 el cual es un indicativo que ambas variables están fuertemente relacionadas uno con la otra.

El análisis bivariado en Google Colab es muy necesario para entender el comportamiento de 2 o más variables categóricas a la vez.

En la figura 3-34, mediante comandos Python se analizó si el nivel académico más alto alcanzado determinaba la implementación e incorporación de TIC en aulas con más frecuencia. Pero fue sorprendente ver que no necesariamente el maestro que tenga el más alto grado académico es aquel que más implementa las TIC en sus enseñanzas, sino al contrario son aquellos maestros que primeramente

fueron Normalistas y luego alcanzaron grado de Licenciatura son los que las implementas las TIC en sus métodos de enseñanza.

```
#Correlacion entre variables categoricas.
tabla_contingencia = pd.crosstab([df['Normalista'],df['Tecnico'],df['Licenciatura'],df['Especialidad'],df['Mestria']],
                                 df['Incorporacion de TIC en aula'])
tabla_contingencia
```

				Incorporacion de TIC en aula	Casi siempre	Nunca	Ocasionalmente	Rara vez	Siempre	
				Mestria						
Normalista		Tecnico	Licenciatura	Especialidad	No	0	1	1	0	0
					No					
					Si	0	4	4	5	2
					Si	2	0	5	1	1
					Si	35	26	74	43	16
					Si	2	0	1	0	1
					Si	1	1	3	5	0
					Si	2	1	4	0	1
					Si	0	1	0	1	0
					No	15	46	56	41	4
					Si	3	0	0	0	0
					Si	0	3	6	1	1
					Si	0	0	0	0	2
					Si	36	30	101	53	10
					Si	5	2	6	5	1
					Si	15	16	43	24	12
					No	0	1	2	2	1
					No	5	4	11	10	1

Figura 3-34: Análisis bivariado de variables categóricas.

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-35, el análisis bivariado entre las variables “internet en colegio” e “Incorporación de TIC en aula” se muestra que en aquellos colegios que no cuentan con internet, son donde en mayor cantidad los maestros logran incorporar las TIC en sus métodos de enseñanza.

		Incorporacion de TIC en aula	Casi siempre	Nunca	Ocasionalmente	Rara vez	Siempre	
		internet en colegio						
		No	69	98	207	139	28	
		Si	52	38	110	52	25	

Figura 3-35: Análisis bivariado de variables cualitativas.

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-36, se muestra de forma clara que aquellos maestros que afirman no tener conocimiento de TIC, son en cantidad mayor a los que afirman implementar TIC. En la opción de implementación “casi siempre” supera la cantidad de maestros que afirman conocer las TIC con 78 maestros, motivo por la cual implementan con más frecuencia. Y en la opción de implementación “Siempre” con una cantidad de maestros de 27 que afirman conocer TIC y también 27 afirman no conocer.

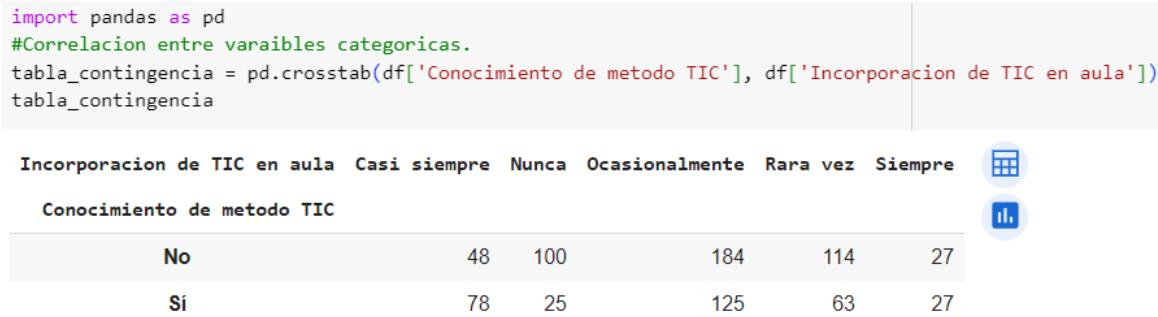


Figura 3-36: Análisis bivariado de variables cualitativas.

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-37, se muestra el análisis de las variables “Requerimiento de internet”, “internet en el colegio” con respecto a la variable “Aula en colegio para uso de computadora”. Donde muestra que aquellos colegios donde se requiere internet, y carecen de este servicio, 252 maestros afirmaron que, si contaban con aulas para uso de computadoras y 119 afirmaron no contar con aulas computacionales, siendo estos valores las cantidades más altas en votación de los maestros.

Claramente el internet es un servicio muy necesario ya que la mayoría de las escuelas la demandan. Pensar en aulas computacionales donde no se cuenta con internet, es decir tener las herramientas, pero no poder usarlas y aprovecharlas al máximo.

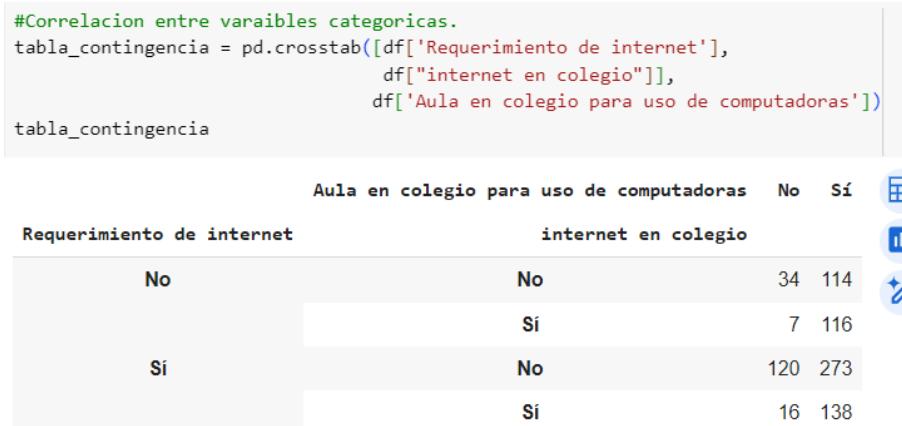


Figura 3-37: Análisis bivariado de variables cualitativas.

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-38, se muestra un análisis bivariado entre las variables de requerimiento de las escuelas y la variable “Aula en colegio para uso de computadoras”. Se observa como 124 maestros afirman si contar con aulas para computadoras en sus escuelas, pero el requerimiento primordial es el internet, también 58 maestros afirman tener aula computacional, pero requieren mobiliario, equipo computacional e internet, seguidamente 80 maestros afirman tener aula computacional, pero requieren mobiliarios e internet.

```
#Correlacion entre variables categoricas.
tabla_contingencia = pd.crosstab([df['Requerimiento de Mobiliario'],df['Requerimiento de equipo computacional'],
                                    df['Requerimiento de internet'], df['internet en colegio']], df['Aula en colegio para uso de computadoras'])
tabla_contingencia
```

			Aula en colegio para uso de computadoras		No	Sí	
			internet en colegio				
Requerimiento de Mobiliario		Requerimiento de equipo computacional	Requerimiento de internet				
No		No	No		7	46	
			Sí		1	37	
				No	19	124	
				Sí	2	43	
		Sí	No		4	12	
			Sí		0	8	
				No	6	11	
				Sí	0	10	
Sí		No	No		20	50	
			Sí		5	64	
				No	44	80	
				Sí	12	47	
		Sí	No		3	6	
			Sí		1	7	
				No	51	58	

**Figura 3-38: Análisis bivariado de requerimientos de los colegios y la existencia de un aula tecnológica.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-39, se observa el análisis bivariado de las variables “Uso de Google académico”, “Uso de Google Drive”, “Incorporación de TIC en aula”, respecto a la variable “Disponibilidad de enseñanza TIC”. Donde la mayor cantidad de maestros con valor de 121 afirman tener la disponibilidad de enseñanza TIC, pero solo lo implementan de forma ocasionalmente y no conocen ni usan Google académico ni Google Drive, y solo 18 maestros tiene la disponibilidad de enseñanza y conocen las herramientas de Google.

```
#Correlacion entre variables categoricas.
tabla_contingencia = pd.crosstab([df['Uso de Google Academic'],df['Uso de Google Drive'],
df['Incorporacion de TIC en aula']], df['Disponibilidad de enseñanza TIC'])
tabla_contingencia
```

		Disponibilidad de enseñanza TIC		No	Sí	
				Incorporacion de TIC en aula		
		No	Casi siempre	5	36	
			Nunca	14	67	
			Ocasionalmente	13	121	
			Rara vez	12	76	
			Siempre	0	14	
		Sí	Casi siempre	0	13	
			Nunca	2	2	
			Ocasionalmente	0	34	
			Rara vez	2	16	
			Siempre	0	5	
		Sí	Casi siempre	1	35	
			Nunca	7	37	
			Ocasionalmente	12	87	
			Rara vez	5	58	
			Siempre	0	18	
		Sí	Casi siempre	2	29	
			Nunca	2	5	
			Ocasionalmente	4	46	
			Rara vez	0	22	
			Siempre	1	15	

**Figura 3-39: Análisis bivariado de variables cualitativas.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-40, se observa un análisis bivariado entre las variables “saber 1”, “Incorporación de TIC en aula” con respecto a la variable “Disponibilidad de enseñanza TIC”. Se puede evidenciar que aquellos maestros que tienen conocimiento del saber 1 e incorporan las TIC de forma casi siempre, rara vez y ocasionalmente, en cantidad son superiores a aquellos maestros que no conocen el saber 1 pese a incorporar las TIC, dando a entender que aquellos maestros que tienen conocimientos en saber 1 son los más dispuestos a enseñar las TIC.

		#Correlacion entre variables categoricas.	
		tabla_contingencia = pd.crosstab([df['Saber 1'], df['Incorporacion de TIC en aula']], df['Disponibilidad de enseñanza TIC'])	
		tabla_contingencia	
Saber 1	Incorporacion de TIC en aula	Disponibilidad de enseñanza TIC	No Sí
Dudoso	Casi siempre	1	7
	Nunca	8	19
	Ocasionalmente	8	30
	Rara vez	4	13
	Siempre	0	1
No	Casi siempre	0	1
	Nunca	1	6
	Ocasionalmente	0	1
	Rara vez	1	5
Si	Casi siempre	7	105
	Nunca	16	86
	Ocasionalmente	21	257
	Rara vez	14	154
	Siempre	1	51

**Figura 3-40:** Análisis bivariado de variables cualitativas.

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-41, se muestra el análisis bivariado entre variables “saber 3”, “Incorporación de TIC en aula” con respecto a la variable “Disponibilidad de enseñanza TIC”. Muestra como 193 maestros afirman que están dispuestos a enseñar las TIC, además son estos los que incorporan de manera ocasionalmente y también conocen saber 3, siendo la cantidad más alta de la encuesta, y los maestros que no tiene conocimiento del saber 3 son los más pocos de la encuesta, pese a que incorporen las TIC.

		#Correlacion entre variables categoricas.	
		tabla_contingencia = pd.crosstab([df['Saber 3'], df['Incorporacion de TIC en aula']], df['Disponibilidad de enseñanza TIC'])	
		tabla_contingencia	
Saber 3	Incorporacion de TIC en aula	Disponibilidad de enseñanza TIC	No Sí
Dudoso	Casi siempre	2	17
	Nunca	10	28
	Ocasionalmente	7	77
	Rara vez	6	46
	Siempre	0	7
No	Casi siempre	0	5
	Nunca	4	25
	Ocasionalmente	7	18
	Rara vez	2	25
Si	Casi siempre	6	91
	Nunca	11	58
	Ocasionalmente	15	193
	Rara vez	11	101
	Siempre	1	45

**Figura 3-41:** Análisis bivariado de variables cualitativas.

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-42, se muestra un análisis entre las variables “saber 5”, “Incorporación de TIC en aula” con respecto a la variable “Disponibilidad de enseñanza TIC”. Dando a conocer que 135 maestros afirman que tiene la disponibilidad de enseñar TIC a sus estudiantes e incorporan de manera ocasionalmente además tiene conocimiento del saber 5, siendo estos maestros la mayor cantidad de la encuesta, mostrando que aquellos maestros que no conocen el saber 5 no tienen la disponibilidad de enseñar TIC a sus estudiantes.

```
#Correlacion entre variables categoricas.
tabla_contingencia = pd.crosstab([df['Saber 5'],
df['Incorporacion de TIC en aula'],
df['Disponibilidad de enseñanza TIC']])
tabla_contingencia
```

		Disponibilidad de enseñanza TIC		No	Sí	
		Incorporacion de TIC en aula				
Dudoso	Casi siempre	1	31			
	Nunca	10	41			
	Ocasionalmente	12	103			
	Rara vez	6	60			
	Siempre	0	6			
No	Casi siempre	2	17			
	Nunca	8	42			
	Ocasionalmente	8	50			
	Rara vez	8	57			
	Siempre	0	8			
Sí	Casi siempre	5	65			
	Nunca	7	28			
	Ocasionalmente	9	135			
	Rara vez	5	55			
	Siempre	1	38			

Figura 3-42: Análisis bivariado de variables cualitativas.

Fuente: Elaboración propia Google Colab, 2023.

### 3.4.3.8. Elección de variables más sobresalientes.

Después de ver y analizar las variables tanto numéricas como categóricas, se determinó que las variables más representativas que serán de utilidad para la creación del modelo de clasificación son “internet en colegio”, “Aula en colegio para uso de computadoras”, “modalidad de clases”, “incorporación de TIC en aula”, “uso de tecnologías mejora el aprendizaje”, llegando así a solo enfocarnos en estas 5 variables.

En la figura 3-43, se muestra la selección de las 5 variables más representativas de todo el conjunto de datos de los maestros evaluados en la encuesta.

```

variables_sobresalientes = ['internet en colegio', 'Aula en colegio para uso de computadoras',
                            'Modalidad de Clases', 'Incorporacion de TIC en aula',
                            'Uso de tecnologias mejoro el aprendizaje']
df1 = df[variables_sobresalientes].copy()

```

**Figura 3-43: Creación de Nuevo DataFrame con variables sobresalientes.****Fuente:** Elaboración propia Google Colab, 2023.

En la figura 3-44, se muestra la creación de la variable “impacto de la inclusión digital”, tomando en cuenta los valores de las 5 variables seleccionadas para representar la clasificación del impacto, aquellos que cuenten con internet, tengan un aula para uso de computadoras, y la modalidad de clases sea diferente a clases frontales o sea clásica, también que la incorporación de TIC sea diferente de nula y que el uso de tecnología haya mejorado el aprendizaje. Para solo tomar en cuenta esta variable se creó un nuevo DataFrame llamado df1 donde se almaceno las 5 variables elegidas y la variable “impacto de la inclusión digital”

```

# Definir una función para evaluar las condiciones y crear la columna 'apto para la inclusion digital'
def evaluar_inclusion_digital(row):
    if (
        row["internet en colegio"] == 'Sí' and
        row['Aula en colegio para uso de computadoras'] == 'Sí' and
        row['Modalidad de Clases'] != "Clases frontales" and
        row["Incorporacion de TIC en aula"] != 'Nunca' and
        row["Uso de tecnologias mejoro el aprendizaje"] == "Sí" #and
        #row["Capacitacion en uso de TIC"] == "Sí"
    ):
        return 'Alto'
    else:
        return 'Bajo'

# Aplicar la función para crear la columna 'apto para la inclusion digital'
df1["impacto de inclusion digital"] = df1.apply(evaluar_inclusion_digital, axis=1)

```

**Figura 3-44: Creación de la variable objetivo.****Fuente:** Elaboración propia Google Colab, 2023.

En la figura 3-45, se muestra de forma visual los valores del DataFrame df1, con las 5 variables representativas y la variable impacto de inclusión digital

	internet en colegio	Aula en colegio para uso de computadoras	Modalidad de Clases	Incorporacion de TIC en aula	Uso de tecnologias mejoro el aprendizaje	impacto de inclusion digital
0	Sí		Sí	Clases frontales	Casi siempre	Sí
1	No		Sí	Clases prácticas	Siempre	Sí
2	Sí		Sí	Clases prácticas	Ocasionalmente	Sí
3	No		Sí	Clases prácticas	Ocasionalmente	Sí
4	No		Sí	Clases prácticas	Ocasionalmente	Sí

**Figura 3-45: Vista del nuevo DataFrame.****Fuente:** Elaboración propia Google Colab, 2023.

### 3.4.3.9. Equilibrio de cantidad de valores.

Para poder realizar un buen modelo de clasificación de machine learning la variable objetivo que en este caso es “impacto de inclusión digital”, debe tener los valores en mismas cantidades, a esto se lo llama equilibrio de valores de la variable objetivo.

En la figura 3-46, mediante código Python en Google Colab, se muestra la cantidad que tienen los valores de la variable objetivo “impacto de la inclusión digital”, siendo 645 los maestros y sus escuelas que están calificadas como de impacto Bajo y solo 173 maestros y sus escuelas tiene una calificación de impacto alto, siendo estos resultados desequilibrio de valores.

```
#mostrar la cantidad de cada valor único en la columna "impacto de inclusion digital"
print(df1["impacto de inclusion digital"].value_counts())

Bajo    645
Alto    173
Name: impacto de inclusion digital, dtype: int64
```

**Figura 3-46: Desequilibrio en variable objetivo.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-47, por medio de códigos Python en Google Colab, se equilibró la variable objetivo “impacto de inclusión digital”. Se observa primeramente los resultados la cantidad que tenían los valores de la variable objetivo, y como después de realizar el equilibrio por medio de código los valores llegaron a ser iguales, pero tomando en cuenta siempre la cantidad más baja de los valores de la variable objetivo, reduciendo así los valores tanto como Alto y Bajo a 173, por lo consiguiente el tamaño de registros de todo el conjunto de datos se quedó en 346.

```
from sklearn.utils import resample

print(df1["impacto de inclusion digital"].value_counts())

# Separar las clases en DataFrames separados
df_class_alto = df1[df1["impacto de inclusion digital"] == 'Alto']
df_class_bajo = df1[df1["impacto de inclusion digital"] == 'Bajo']

# Determinar la cantidad de muestras deseada para equilibrar las clases
target_sample_size = min(len(df_class_alto), len(df_class_bajo))

# Submuestrear cada clase para igualar la cantidad de muestras
df_class_alto_downsampled = resample(df_class_alto, replace=False, n_samples=target_sample_size, random_state=42)
df_class_bajo_downsampled = resample(df_class_bajo, replace=False, n_samples=target_sample_size, random_state=42)

# Combinar las clases balanceadas en un nuevo DataFrame
df2_balanced = pd.concat([df_class_alto_downsampled, df_class_bajo_downsampled])

# Verificar el equilibrio de clases en el nuevo DataFrame
print(df2_balanced["impacto de inclusion digital"].value_counts())

Bajo    645
Alto    173
Name: impacto de inclusion digital, dtype: int64
Alto    173
Bajo    173
Name: impacto de inclusion digital, dtype: int64
```

**Figura 3-47: Equilibrio de la variable objetivo.**

Fuente: Elaboración propia Google Colab, 2023.

Después de realizar el proceso que muestra la figura 3-47 el conjunto de datos se almaceno en el DataFrame df2\_balanced, donde los registros pasaron a ser 346 y las variables 6.

### 3.4.3.10. One Hot Encoding:

En la figura 3-48, se muestra las primeras variables cualitativas ya convertidas por One Hot Encoding a variables con valores entre 0 y 1, donde 0 es igual al valor No y 1 es igual al valor Si; que están listas para usar en nuestro modelo de clasificación.

```
# Aplicar One-Hot Encoding a todas las columnas categóricas en df1
df_encoded = pd.get_dummies(df2_balanced)

# Ver el DataFrame con One-Hot Encoding
df_encoded.head()
```

	internet en colegio_No	internet en colegio_Sí	Aula en colegio para uso de computadoras_No	Aula en colegio para uso de computadoras_Sí	Modalidad de clases_Classes de laboratorios	Modalidad de Clases_Classes frontales	Modalidad de Clases_Classes prácticas	Modalidad de Clases_Seminarios	Modalidad de Clases_Talleres
851	0	1	0	1	0	0	1	0	0
286	0	1	0	1	0	0	1	0	0
502	0	1	0	1	0	0	1	0	0
397	0	1	0	1	0	0	1	0	0
687	0	1	0	1	0	0	1	0	0

**Figura 3-48: Conjunto de Datos listo para el Modelo.**

Fuente: Elaboración propia Google Colab, 2023.

Llegado a este punto se tiene un conjunto de datos ya analizado, y hecho un tratamiento correctivo y se encuentra listo para poder hacer el modelado de este conjunto de datos.

### 3.4.4. Modelado.

En la etapa de modelamiento de machine learning como aprendizaje supervisado, es la etapa donde se realiza el modelamiento en los algoritmos de clasificación más conocidos como árboles de decisión, máquinas de vectores de soporte (SVM), clasificación de Naive Bayes, K vecinos más cercanos (KNN), regresión logística y redes neuronales, con el objeto de ver el comportamiento del modelo, y así poder determinar el mejor entre ellos.

#### 3.4.4.1. Regresión Logística.

Aunque a menudo se utiliza para problemas de regresión, también se puede utilizar para problemas de clasificación binaria.

En la figura 3-49, se muestra mediante código Python en Google Colab y con la ayuda de las librerías de Sklearn, la creación del modelo de clasificación con el algoritmo Regresión Logística. Para esto tomando en cuenta los valores de las variables de estudio, seleccionando así las variables predictoras y la variable objetivo, seguido de la división de datos en conjunto de entrenamiento y conjunto de prueba “x\_train,

x\_test, y\_train, y\_test”, para luego realizar la creación del modelo “model”, seguidamente pasando por el entrenamiento para que el modelo pueda aprender a clasificar y por último la realización de la prueba.

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
import matplotlib.pyplot as plt

# Variables predictoras
X = df_encoded[['internet en colegio_Sí', 'Aula en colegio para uso de computadoras_Sí',
                 'Modalidad de Clases_Clases de laboratorios','Modalidad de Clases_Clases prácticas',
                 'Modalidad de Clases_Clases prácticas','Incorporacion de TIC en aula_Casi siempre',
                 'Incorporacion de TIC en aula_Ocasionalmente','Incorporacion de TIC en aula_Rara vez',
                 'Incorporacion de TIC en aula_Siempre','Uso de tecnologias mejoro el aprendizaje_Sí',]]

# Variable objetivo
y = df_encoded['impacto de inclusion digital_Alto']

# Dividision de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Creacion del modelo de regresión logística
model = LogisticRegression(max_iter=1000)

#Entrenamiento del modelo
model.fit(X_train, y_train)

#Prueba del modelo
y_pred = model.predict(X_test)
```

**Figura 3-49: Preparación, entrenamiento y prueba del Modelo en Regresión Logística Binaria.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-50, mediante códigos Python en Google Colab muestra el cálculo de las métricas de evaluación del modelo de clasificación de Regresión Logística. Los valores de las métricas del modelo indica un buen aprendizaje y que es capaz de clasificar el impacto de inclusión digital en las escuelas públicas de Bolivia.

```
# Calculo de las métricas de evaluación
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
y_prob = model.predict_proba(X_test)[:, 1]
roc_auc = roc_auc_score(y_test, y_prob)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred)

# Impresion de las métricas
print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1-score:", f1)
print("AUC-ROC:", roc_auc)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)
```

**Figura 3-50: Métricas de evaluación y matriz de confusión del modelo en Regresión Logística Binaria.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-51 se muestra los resultados de las métricas de evaluación como Accuracy con valor 0,97; Precision con valor 0,95; Recall con valor 1,0; F1\_score con valor 0,97; y AUC\_ROC con valor 0,99; también los valores de la matriz de confusión como TP = 30, FP = 2, FN = 0, TN = 38.

```
# Curva ROC
fpr, tpr, thresholds = roc_curve(y_test, y_prob)
plt.figure()
plt.plot(fpr, tpr, color='darkorange', lw=2, label='Curva ROC (AUC = %0.2f)' % roc_auc)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de Falsos Positivos')
plt.ylabel('Tasa de Verdaderos Positivos')
plt.title('Curva ROC')
plt.legend(loc="lower right")
plt.show()

Accuracy: 0.9714285714285714
Precision: 0.95
Recall: 1.0
F1-score: 0.9743589743589743
AUC-ROC: 0.9971217105263158

Matriz de Confusión:
[[30  2]
 [ 0 38]]
```

**Figura 3-51: Valores de métricas de evaluación del modelo de Regresión Logística Binaria.**

Fuente: Elaboración propia Google Colab, 2023.

#### 3.4.4.2. Máquinas de Vectores de Soporte (SVM)

En la figura 3-52, muestra mediante código Python en Google Colab y con la ayuda de las librerías de Sklearn, la creación del modelo de clasificación con el algoritmo Maquinas de vectores de soporte (SVM). Para esto tomando en cuenta los valores de las variables de estudio, seleccionando así las variables predictoras y la variable objetivo, seguido de la división de datos en conjunto de entrenamiento y conjunto de prueba “x\_train, x\_test, y\_train, y\_test”, para luego realizar la creación del modelo “model”, seguidamente pasando por el entrenamiento para que el modelo pueda aprender a clasificar y por último la realización de la prueba.

```
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
import matplotlib.pyplot as plt

#Variables predictoras
X = df_encoded[["internet en colegio_Sí", "Aula en colegio para uso de computadoras_Sí",
                 "Modalidad de Clases_Clases de laboratorios", "Modalidad de Clases_Clases prácticas",
                 'Modalidad de Clases_Clases prácticas', 'Incorporacion de TIC en aula_Casi siempre',
                 'Incorporacion de TIC en aula_Ocasionalmente', 'Incorporacion de TIC en aula_Rara vez',
                 'Incorporacion de TIC en aula_Siempre', 'Uso de tecnologias mejoro el aprendizaje_Sí', ]]

#Variable objetivo
y = df_encoded["impacto de inclusion digital_Alto"]

# Division de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Creacion del modelo SVM con kernel radial (RBF)
svm_model = SVC(kernel='rbf', random_state=42, probability=True)

#Entrenamiento del modelo
svm_model.fit(X_train, y_train)

#Predicciones del conjunto de prueba
y_pred_svm = svm_model.predict(X_test)
```

**Figura 3-52: Modelo en Maquinas de Vectores de Soporte (SVM).**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-53, mediante códigos Python en Google Colab muestra el cálculo de las métricas de evaluación del modelo de clasificación de Maquinas de vectores de soporte (SVM). Los valores de las métricas del modelo indica un buen aprendizaje y que es capaz de clasificar el impacto de inclusión digital en las escuelas públicas de Bolivia.

```
# Calculo de métricas de evaluación
accuracy_svm = accuracy_score(y_test, y_pred_svm)
precision_svm = precision_score(y_test, y_pred_svm)
recall_svm = recall_score(y_test, y_pred_svm)
f1_svm = f1_score(y_test, y_pred_svm)
y_prob_svm = svm_model.predict_proba(X_test)[:, 1]
roc_auc_svm = roc_auc_score(y_test, y_prob_svm)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred_svm)

# Impresion de las métricas
print("Accuracy (SVM):", accuracy_svm)
print("Precision (SVM):", precision_svm)
print("Recall (SVM):", recall_svm)
print("F1-score (SVM):", f1_svm)
print("AUC-ROC (SVM):", roc_auc_svm)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)
```

**Figura 3-53: Resultado del modelo en Maquinas de Vectores de Soporte (SVM).**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-54, se muestra los resultados de las métricas de evaluación como Accuracy con valor 0,97; Precision con valor 0,95; Recall con valor 1,0; F1\_score con valor 0,97; y AUC\_ROC con valor 0,99; también los valores de la matriz de confusión como TP = 28, FP = 4, FN = 0, TN = 38.

```
# Curva ROC
fpr_svm, tpr_svm, thresholds_svm = roc_curve(y_test, y_prob_svm)
plt.figure()
plt.plot(fpr_svm, tpr_svm, color='darkorange', lw=2, label='Curva ROC (AUC = %0.2f)' % roc_auc_svm)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de Falsos Positivos')
plt.ylabel('Tasa de Verdaderos Positivos')
plt.title('Curva ROC (SVM)')
plt.legend(loc="lower right")
plt.show()

Accuracy (SVM): 0.9428571428571428
Precision (SVM): 0.9047619047619048
Recall (SVM): 1.0
F1-score (SVM): 0.9500000000000001
AUC-ROC (SVM): 0.989720394736842

Matriz de Confusión:
[[28  4]
 [ 0 38]]
```

**Figura 3-54: Valores de métricas de evaluación del modelo Maquinas de Vectores de Soporte (SVM).**

Fuente: Elaboración propia Google Colab, 2023.

### 3.4.4.3. K Vecinos más Cercanos (K-NN).

En la figura 3-55, muestra mediante código Python en Google Colab y con la ayuda de las librerías de Sklearn, la creación del modelo de clasificación con el algoritmo K Vecinos más Cercanos (K-NN). Para esto tomando en cuenta los valores de las variables de estudio, seleccionando así las variables predictoras y la variable objetivo, seguido de la división de datos en conjunto de entrenamiento y conjunto de prueba “x\_train, x\_test, y\_train, y\_test”, para luego realizar la creación del modelo “model”, seguidamente pasando por el entrenamiento para que el modelo pueda aprender a clasificar y por último la realización de la prueba.

```
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
import matplotlib.pyplot as plt

#Variables predictoras
X = df_encoded[["internet en colegio_Sí", "Aula en colegio para uso de computadoras_Sí",
                 "Modalidad de Clases_Clases de laboratorios", 'Modalidad de Clases_Clases prácticas',
                 'Modalidad de Clases_Clases prácticas', 'Incorporacion de TIC en aula_Casi siempre',
                 'Incorporacion de TIC en aula_Ocasionalmente', 'Incorporacion de TIC en aula_Rara vez',
                 'Incorporacion de TIC en aula_Siempre', 'Uso de tecnologias mejoro el aprendizaje_Sí',]]

#Varaible objetivo
y = df_encoded["impacto de inclusion digital_Alto"]

# Dividision de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Creacion del modelo de k-Nearest Neighbors (k-NN)
knn_model = KNeighborsClassifier(n_neighbors=3)

#Entrenamiento del modelo
knn_model.fit(X_train, y_train)

#Predicciones del conjunto de prueba
y_pred_knn = knn_model.predict(X_test)
```

**Figura 3-55: Modelo en K Vecinos más cercanos (K-NN).**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-56, mediante códigos Python en Google Colab muestra el cálculo de las métricas de evaluación del modelo de clasificación de K Vecinos más Cercanos (K-NN). Regresión Logística como Accuracy con valor 0,88; Precision con valor 0,85; Recall con valor 0,94; F1\_score con valor 0,90. Los valores de las métricas del modelo indica un buen aprendizaje y que es capaz de clasificar el impacto de inclusión digital en las escuelas públicas de Bolivia. También los valores de la matriz de confusión como TP = 26, FP = 6, FN = 2, TN = 36.

```

# Calculo de métricas de evaluación
accuracy_knn = accuracy_score(y_test, y_pred_knn)
precision_knn = precision_score(y_test, y_pred_knn)
recall_knn = recall_score(y_test, y_pred_knn)
f1_knn = f1_score(y_test, y_pred_knn)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred_knn)

# Impresión de las métricas
print("Accuracy (k-NN):", accuracy_knn)
print("Precision (k-NN):", precision_knn)
print("Recall (k-NN):", recall_knn)
print("F1-score (k-NN):", f1_knn)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)

# No se calcula AUC-ROC para k-NN ya que no genera probabilidades directas

# Curva ROC (No aplicable para k-NN)
# k-NN no genera probabilidades directas, por lo que no se puede calcular una curva ROC en este caso.

```

Accuracy (k-NN): 0.8857142857142857  
 Precision (k-NN): 0.8571428571428571  
 Recall (k-NN): 0.9473684210526315  
 F1-score (k-NN): 0.9

 Matriz de Confusión:  
 [[26 6]  
 [ 2 36]]

**Figura 3-56: Resultados del modelo en K Vecinos más cercanos (K-NN).**

Fuente: Elaboración propia Google Colab, 2023.

#### 3.4.4.4. Árboles de Decisión

En la figura 3-57, muestra mediante código Python en Google Colab y con la ayuda de las librerías de Sklearn, la creación del modelo de clasificación con el algoritmo Árboles de Decisión. Para esto tomando en cuenta los valores de las variables de estudio, seleccionando así las variables predictoras y la variable objetivo, seguido de la división de datos en conjunto de entrenamiento y conjunto de prueba “x\_train, x\_test, y\_train, y\_test”, para luego realizar la creación del modelo “model”, seguidamente pasando por el entrenamiento para que el modelo pueda aprender a clasificar y por último la realización de la prueba.

```
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
import matplotlib.pyplot as plt

#Variable predictora
X = df_encoded[['internet en colegio_Sí', 'Aula en colegio para uso de computadoras_Sí',
                 'Modalidad de Clases_Clases de laboratorios','Modalidad de Clases_Clases prácticas',
                 'Modalidad de Clases_Clases prácticas','Incorporacion de TIC en aula_Casi siempre',
                 'Incorporacion de TIC en aula_Ocasionalmente','Incorporacion de TIC en aula_Rara vez',
                 'Incorporacion de TIC en aula_Siempre','Uso de tecnologias mejoro el aprendizaje_Sí',]]

#Variable objetivo
y = df_encoded[['impacto de inclusion digital_Alto']]

# División de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Crear un modelo de Árboles de Decisión
tree_model = DecisionTreeClassifier(random_state=42)

#Entrenamiento del modelo
tree_model.fit(X_train, y_train)

#Predicciones en el conjunto de prueba
y_prob_tree = tree_model.predict_proba(X_test)[:, 1]
y_pred_tree = (y_prob_tree > 0.5).astype(int)
```

**Figura 3-57: Modelo en Árbol de Decisiones.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-58, mediante códigos Python en Google Colab muestra el cálculo de las métricas de evaluación del modelo de clasificación de Árboles de Decisión. Los valores de las métricas del modelo indica un buen aprendizaje y que es capaz de clasificar el impacto de inclusión digital en las escuelas públicas de Bolivia.

```
# Calculo de métricas de evaluación
accuracy_tree = accuracy_score(y_test, y_pred_tree)
precision_tree = precision_score(y_test, y_pred_tree)
recall_tree = recall_score(y_test, y_pred_tree)
f1_tree = f1_score(y_test, y_pred_tree)
roc_auc_tree = roc_auc_score(y_test, y_prob_tree)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred_tree)

# Impresión de las métricas
print("Accuracy (Decision Tree):", accuracy_tree)
print("Precision (Decision Tree):", precision_tree)
print("Recall (Decision Tree):", recall_tree)
print("F1-score (Decision Tree):", f1_tree)
print("AUC-ROC (Decision Tree):", roc_auc_tree)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)
```

**Figura 3-58: Resultados del modelo en Árbol de Decisiones.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-59, mediante códigos Python en Google Colab muestra el cálculo de las métricas de evaluación del modelo de clasificación con algoritmo de árbol de decisiones. Regresión Logística como Accuracy con valor 0,95; Precision con valor 0,97; Recall con valor 0,94; F1\_score con valor 0,95; AUC-ROC con valor 0,97. Los valores de las métricas del modelo indica un buen aprendizaje y que es capaz de clasificar el impacto de inclusión digital en las escuelas públicas de Bolivia. También los valores de la matriz de confusión como TP = 31, FP = 1, FN = 2, TN = 36.

```
# Curva ROC
fpr_tree, tpr_tree, thresholds_tree = roc_curve(y_test, y_prob_tree)
plt.figure()
plt.plot(fpr_tree, tpr_tree, color='darkorange', lw=2, label='Curva ROC (AUC = %0.2f)' % roc_auc_tree)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de Falsos Positivos')
plt.ylabel('Tasa de Verdaderos Positivos')
plt.title('Curva ROC (Decision Tree)')
plt.legend(loc="lower right")
plt.show()

Accuracy (Decision Tree): 0.9571428571428572
Precision (Decision Tree): 0.972972972972973
Recall (Decision Tree): 0.9473684210526315
F1-score (Decision Tree): 0.9599999999999999
AUC-ROC (Decision Tree): 0.9708059210526316

Matriz de Confusión:
[[31  1]
 [ 2 36]]
```

**Figura 3-59: Valores de métricas de evaluación del modelo árbol de decisiones.**

Fuente: Elaboración propia Google Colab, 2023.

#### 3.4.4.5. Naive Bayes.

En la figura 3-60, muestra mediante código Python en Google Colab y con la ayuda de las librerías de Sklearn, la creación del modelo de clasificación con el algoritmo Naive Bayes. Para esto tomando en cuenta los valores de las variables de estudio, seleccionando así las variables predictoras y la variable objetivo, seguido de la división de datos en conjunto de entrenamiento y conjunto de prueba “x\_train, x\_test, y\_train, y\_test”, para luego realizar la creación del modelo “model”, seguidamente pasando por el entrenamiento para que el modelo pueda aprender a clasificar y por último la realización de la prueba.

```
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
import matplotlib.pyplot as plt

# Variables predictoras
X = df_encoded[['internet en colegio_Sí', 'Aula en colegio para uso de computadoras_Sí',
                 'Modalidad de Clases_Clases de laboratorios','Modalidad de Clases_Clases prácticas',
                 'Modalidad de Clases_Clases prácticas','Incorporacion de TIC en aula_Casi siempre',
                 'Incorporacion de TIC en aula_Ocasionalmente','Incorporacion de TIC en aula_Rara vez',
                 'Incorporacion de TIC en aula_Siempre','Uso de tecnologias mejoró el aprendizaje_Sí',]]

# Variable objetivo
y = df_encoded['impacto de inclusión digital_Alto']

# División de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Crear del Modelo
nb_model = GaussianNB()

#Entrenamiento del modelo
nb_model.fit(X_train, y_train)

#Predicciones en el conjunto de prueba
y_pred_nb = nb_model.predict(X_test)
```

**Figura 3-60: Modelo en Naive Bayes.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-61, mediante códigos Python en Google Colab muestra el cálculo de las métricas de evaluación del modelo de clasificación de Naive Bayes. Los valores de las métricas del modelo indican un buen aprendizaje y que es capaz de clasificar el impacto de inclusión digital en las escuelas públicas de Bolivia.

```
# Cálculo de métricas de evaluación
accuracy_nb = accuracy_score(y_test, y_pred_nb)
precision_nb = precision_score(y_test, y_pred_nb)
recall_nb = recall_score(y_test, y_pred_nb)
f1_nb = f1_score(y_test, y_pred_nb)
y_prob_nb = nb_model.predict_proba(X_test)[:, 1]
roc_auc_nb = roc_auc_score(y_test, y_prob_nb)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred_nb)

# Impresión de métricas
print("Accuracy (Naive Bayes):", accuracy_nb)
print("Precision (Naive Bayes):", precision_nb)
print("Recall (Naive Bayes):", recall_nb)
print("F1-score (Naive Bayes):", f1_nb)
print("AUC-ROC (Naive Bayes):", roc_auc_nb)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)
```

**Figura 3-61: Resultado del modelo en Naive Bayes.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-62, mediante códigos Python en Google Colab muestra el cálculo de las métricas de evaluación del modelo de clasificación de Naive Bayes, como Accuracy con valor 0,97; Precision con valor 0,95; Recall con valor 1,0; F1\_score con valor 0,97; AUC-ROC con valor 0,98. Los valores de las métricas del modelo indica un buen aprendizaje y que es capaz de clasificar el impacto de inclusión digital en las escuelas públicas de Bolivia. También los valores de la matriz de confusión como TP = 30, FP = 2, FN = 0, TN = 38.

```
# Curva ROC
fpr_nb, tpr_nb, thresholds_nb = roc_curve(y_test, y_prob_nb)
plt.figure()
plt.plot(fpr_nb, tpr_nb, color='darkorange', lw=2, label='Curva ROC (AUC = %0.2f)' % roc_auc_nb)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de Falsos Positivos')
plt.ylabel('Tasa de Verdaderos Positivos')
plt.title('Curva ROC (Naive Bayes)')
plt.legend(loc="lower right")
plt.show()

Accuracy (Naive Bayes): 0.9714285714285714
Precision (Naive Bayes): 0.95
Recall (Naive Bayes): 1.0
F1-score (Naive Bayes): 0.9743589743589743
AUC-ROC (Naive Bayes): 0.9814967105263158

Matriz de Confusión:
[[30  2]
 [ 0 38]]
```

**Figura 3-62: Valores de métricas de evaluación del modelo Naive Bayes.**

Fuente: Elaboración propia Google Colab, 2023.

#### 3.4.4.6. Redes Neuronales.

En la figura 3-63, se muestra que mediante código Python en Google Colab y con la ayuda de las librerías de Sklearn, la creación del modelo de clasificación con el algoritmo Redes Neuronales. Para esto tomando en cuenta los valores de las variables de estudio, seleccionando así las variables predictoras, almacenadas en el DataFrame “X” y la variable objetivo almacenada en el DataFrame “y”, seguido de la división de datos en conjunto de entrenamiento en las variables ““x\_train, y\_train” y conjunto de prueba en las variables “x\_test, y\_test”, para luego realizar el escalamiento de las características normalizandolas para que tengan una distribución uniforme, seguido la creación del modelo “model”, seguidamente pasando por el entrenamiento para que el modelo pueda aprender a clasificar y por último la realización de la prueba.

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from tensorflow import keras
from tensorflow.keras import layers
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt

#Variables predictoras
X = df_encoded[['Departamento','internet en colegio_Sí', "Aula en colegio para uso de computadoras_Sí",
                 "Modalidad de Clases_Clases de laboratorios",'Modalidad de Clases_Clases prácticas',
                 'Incorporacion de TIC en aula_Casi siempre','Incorporacion de TIC en aula_Ocasionalmente',
                 'Incorporacion de TIC en aula_Rara vez','Incorporacion de TIC en aula_Siempre',
                 'Uso de tecnologias mejoro el aprendizaje_Sí']]

#Variables objetivo
y = df_encoded["impacto de inclusion digital_Alto"]

# Division de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Escalamiento de las características para mejorar el rendimiento de la red neuronal
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Creacion de un modelo de red neuronal artificial (ANN)
model = keras.Sequential([
    layers.Input(shape=(X_train_scaled.shape[1],)), # Capa de entrada con el número de características
    layers.Dense(64, activation='relu'), # Capa oculta con 64 neuronas y función de activación ReLU
    layers.Dense(1, activation='sigmoid') # Capa de salida con 1 neurona y función de activación sigmoide (p)
])

# Compilacion del modelo
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Entrenamiento el modelo
model.fit(X_train_scaled, y_train, epochs=100, batch_size=64, validation_split=0.2)

#Predicciones en el conjunto de prueba
y_prob_ann = model.predict(X_test_scaled)
y_pred_ann = (y_prob_ann > 0.5).astype(int)
```

Figura 3-63: Modelo en Redes Neuronales profundas (ANN).

Fuente: Elaboración propia Google Colab, 2023.

La figura 3-64, generada utilizando códigos Python en Google Colab muestra el cálculo de las métricas de evaluación del modelo de clasificación de Redes Neuronales. Los valores de las métricas del modelo indica el grado de aprendizaje y la capacidad de clasificar el impacto de inclusión digital en las escuelas públicas de Bolivia.

```

# Calculo de métricas de evaluación
accuracy_ann = accuracy_score(y_test, y_pred_ann)
precision_ann = precision_score(y_test, y_pred_ann)
recall_ann = recall_score(y_test, y_pred_ann)
f1_ann = f1_score(y_test, y_pred_ann)
roc_auc_ann = roc_auc_score(y_test, y_prob_ann)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred_ann)

# Impresion de las métricas
print("Accuracy (ANN):", accuracy_ann)
print("Precision (ANN):", precision_ann)
print("Recall (ANN):", recall_ann)
print("F1-score (ANN):", f1_ann)
print("AUC-ROC (ANN):", roc_auc_ann)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)

```

**Figura 3-64: Resultado del modelo en Redes Neuronales profunda (ANN).****Fuente:** Elaboración propia Google Colab, 2023.

En la figura 3-65, mediante códigos Python en Google Colab muestra el cálculo de las métricas de evaluación del modelo de clasificación de Naive Bayes. Regresión Logística como Accuracy con valor 0,98; Precision con valor 0,97; Recall con valor 1,0; F1\_score con valor 0,98; AUC-ROC con valor 0,99. Los valores de las métricas del modelo indica un buen aprendizaje y que es capaz de clasificar el impacto de inclusión digital en las escuelas públicas de Bolivia. También los valores de la matriz de confusión como TP = 31, FP = 1, FN = 0, TN = 38.

```

# Curva ROC
fpr_ann, tpr_ann, thresholds_ann = roc_curve(y_test, y_prob_ann)
plt.figure()
plt.plot(fpr_ann, tpr_ann, color='darkorange', lw=2, label='Curva ROC (AUC = %0.2f)' % roc_auc_ann)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de Falsos Positivos')
plt.ylabel('Tasa de Verdaderos Positivos')
plt.title('Curva ROC (ANN)')
plt.legend(loc="lower right")
plt.show()

```

```

Accuracy (ANN): 0.9857142857142858
Precision (ANN): 0.9743589743589743
Recall (ANN): 1.0
F1-score (ANN): 0.9870129870129869
AUC-ROC (ANN): 0.9979440789473684

```

```

Matriz de Confusión:
[[31  1]
 [ 0 38]]

```

**Figura 3-65: Valores de métricas de evaluación del modelo Redes Neuronales (ANN).****Fuente:** Elaboración propia Google Colab, 2023.

#### 3.4.4.7. Guardar modelo.

Una vez realizada el entrenamiento y evaluación del modelo mediante sus métricas, es necesario realizar la exportación del modelo, para que este sea utilizado en otro conjunto de datos o donde se lo requiera tal y como se observa en la figura 3-66.

```
import joblib  
  
# Guardar el modelo  
joblib.dump(tree_model, 'Modelo1.pkl')
```

Figura 3-66: Exportación del modelo entrenado.

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-66, se muestra que para poder guardar o exportar un modelo se hace uso de la librería joblib y mediante el comando dump lo podemos guardar, para que sea utilizado por otros conjuntos de datos.

#### 3.4.4.8. Cargar modelo.

Para que cuando el modelo sea requerido para usarlo en otro conjunto de datos, se debe cargar tal como se muestra en la figura 3-67.

```
#Caragar modelo ya entrenado.  
import os  
import joblib  
  
# Obtén la ruta completa al archivo Modelo1.pkl  
ruta_modelo = os.path.join(os.getcwd(), 'Modelo1.pkl')  
  
# Cargar el modelo  
modelo_entrenado = joblib.load('/content/drive/MyDrive/Modelo1.pkl')
```

Figura 3-67: Importación del modelo ya entrenado.

Fuente: Elaboración propia Google Colab, 2023.

Para poder usar el modelo ya cargado con otro conjunto de datos, se debe primeramente cargar el nuevo conjunto de datos en figura3-68.

```
df_nuevo = pd.read_excel('/content/drive/MyDrive/df_encoded4.xlsx')  
df_nuevo
```

Figura 3-68: DataFrame del nuevo conjunto de datos.

Fuente: Elaboración propia Google Colab, 2023.

Para poder usar el modelo con el nuevo conjunto de datos se debe usar el comando predict tal como se ve en la figura 3-69. Donde los resultados predichos se encuentran en la variable nuevas\_predicciones de tipo arreglo.

```
nuevas_predicciones = modelo_entrenado.predict(df_nuevo)
nuevas_predicciones

array([0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1], dtype=uint8)
```

**Figura 3-69: Uso del modelo con el nuevo conjunto de datos.**

Fuente: Elaboración propia Google Colab, 2023.

Luego de contar con las predicciones se las realiza la unión al conjunto de datos nuevo, tal como se observa en la figura 3-70.

df_nuevo.insert(0,'Prediccion',nuevas_predicciones,allow_duplicates=False)												
df_nuevo.head(300)												
Prediccion	internet_en_colegio_Sí	Aula_en_colegio_para_uso_de_computadoras_Sí	Modalidad_de_Clases_Clases_de_laboratorios	Modalidad_de_Clases_Clases_prácticas	Modalidad_de_Clases_Clases_prácticas	Incorporacion_de_TIC_en_aula_Casi_siempre	Incorporacion_de_TIC_en_aula_Ocasionalmente	Incorporacion_de_TIC_en_aula_Rara_vez	Incorporacion_de_TIC_en_aula_Siempre	Incorporacion_de_TIC_en_aula_Siempre	Uso_de_tecnologías_mejoro_el_aprendizaje_Sí	
0	0	0	1	0	1	1	0	1	0	0	1	
1	0	0	0	1	0	0	0	1	0	0	0	
2	0	0	0	0	1	1	0	1	0	0	0	
3	0	0	0	0	1	1	0	0	0	0	0	
4	0	1	1	0	1	1	0	0	0	0	1	
...	...	...	...	...	...	...	...	...	...	...	...	
56	1	1	1	0	0	0	1	0	0	0	1	
57	1	1	1	0	1	1	0	0	1	0	1	
58	1	1	1	0	1	1	0	0	1	0	1	
59	1	1	1	0	1	1	1	0	0	0	1	
60	1	1	1	0	0	0	0	1	0	0	1	

**Figura 3-70: Unión de predicciones y conjunto de datos nuevos.**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 3-70 se observa como el modelo exportado que luego fue cargado para ser usado con otro conjunto de dato

## 4. Resultados y Discusión

Con los datos obtenidos y analizados, se buscó determinar un algoritmo que sea capaz de clasificar el tipo de impacto que tuvo el programa de inclusión digital en las escuelas públicas de Bolivia. Para esto se usó datos de encuestas realizadas a maestros de varios colegios de todos los departamentos de Bolivia. Con la ayuda de códigos en programación en Python y el servicio de Jupyter Notebook que se hospeda en Google llamado Google Colab, se pudo realizar la parte de Análisis y exploración de datos, con el cual se pudo obtener valiosa información. También ayudo a realizar el modelo con los algoritmo de clasificación de aprendizaje supervisado de machine learning el cual ayuda a la clasificación del impacto de inclusión digital.

### 4.1 Resultados de la exploración y análisis del conjunto de datos.

Al analizar los datos se dio a conocer que los maestros de las escuelas públicas, no se quedan estáticos en cuestión de nivel académico, y se puede notar que los maestros buscan mejorar su nivel académico y alcanzar mejores oportunidades. Los resultados más altos en nivel académico de los maestros son el 73,4% en Licenciatura y 69,3% en Normalista. Siendo pocos, pero significativos se tiene 17,2% en nivel de Especialidad y 5,3% en Maestría, demostrando que los maestros tienen el interés de mejorar su nivel académico cada vez más, como se puede observar en la siguiente figura 4-1.

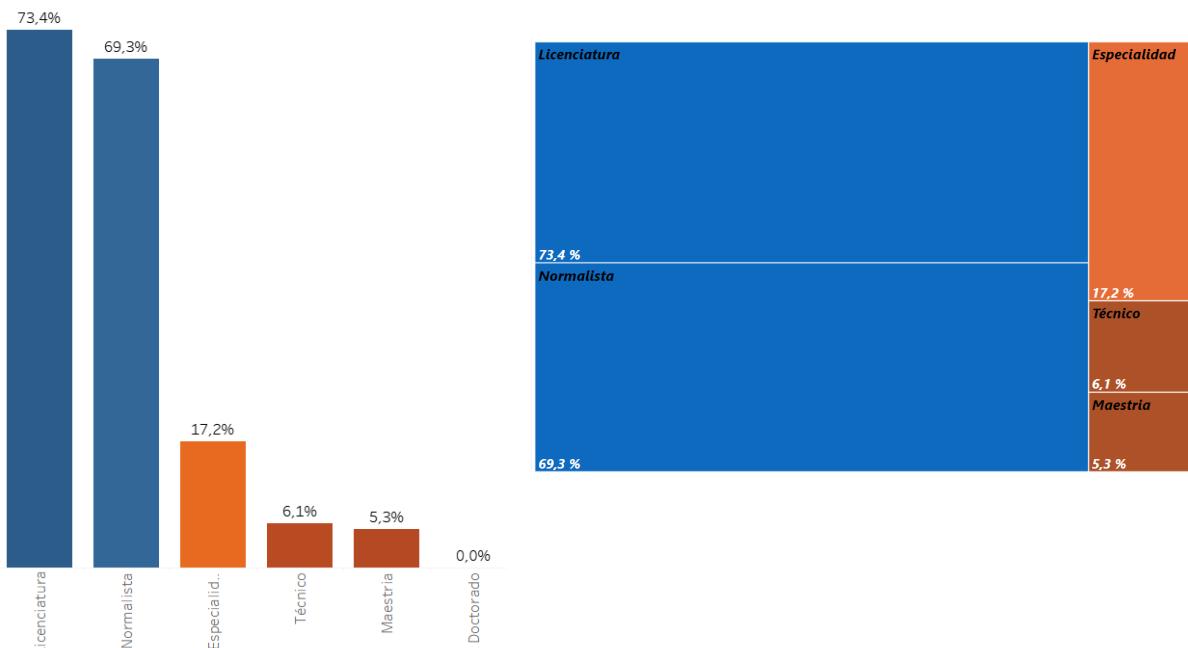
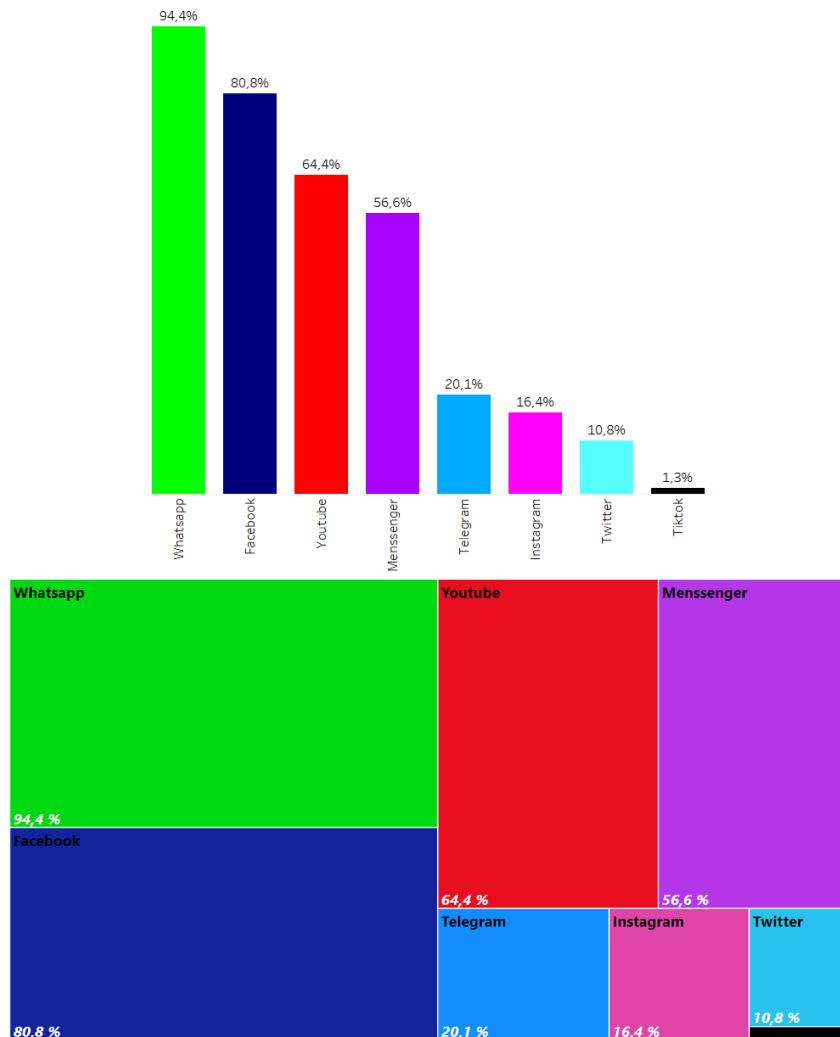


Figura 4-1: Nivel académico de los maestros en las escuelas públicas.

Fuente: Elaboración propia Tableau, 2023.

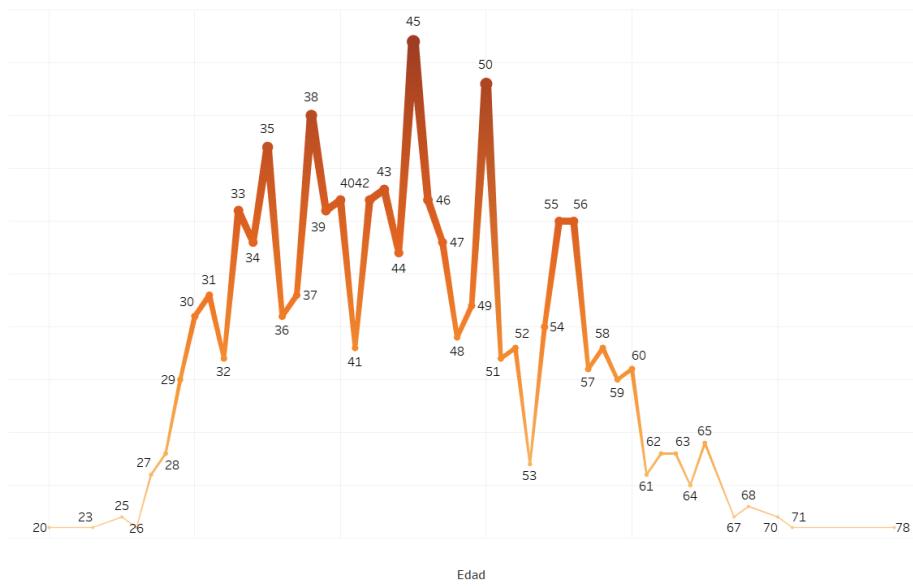
En la figura 4-2, se muestra que los maestros de las escuelas públicas de Bolivia tienen conocimientos en redes sociales, y no solo conocimiento, sino que usan redes sociales, dándonos a entender que también van de la mano de las nuevas tecnologías de comunicación que van saliendo con el avance de la tecnología. La red social más conocida y usada por los maestros es WhatsApp con 94,4% de frecuencia, seguidos por Facebook, YouTube, Messenger, y demás, lo cual da a conocer que se usan estas herramientas como medios de ayuda para la educación.



**Figura 4-2: Uso de redes sociales en los maestros de las escuelas públicas.**

Fuente: Elaboración propia Tableau, 2023.

En la figura 4-3 muestra que, en cuestión de redes sociales se observó que la edad no es limitante para conocer y usarlas, ya que en su mayoría los maestros se encuentran en edades de 30 a 60 años. Como la figura 4-2 indica que los maestros son conocedores y usan redes sociales y en la figura 4-3 muestra que la mayoría de los encuestados son de edades mayores, podemos entender que los maestros no se limitan ante la tecnología pese a sus edades.

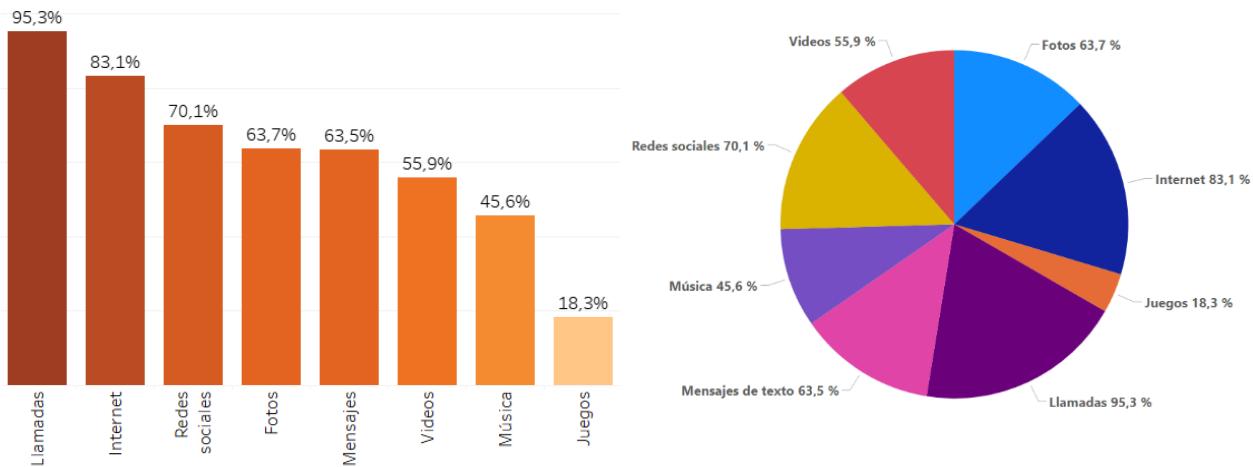


**Figura 4-3: Resultado de mayoría de edad en los maestros.**

Fuente: Elaboración propia Tableau, 2023.

El uso de herramientas como computadoras, tablets, celulares, el internet y demás son muy importantes en la educación, y ante esto se encontró resultados del uso que le dan los maestros a estas herramientas.

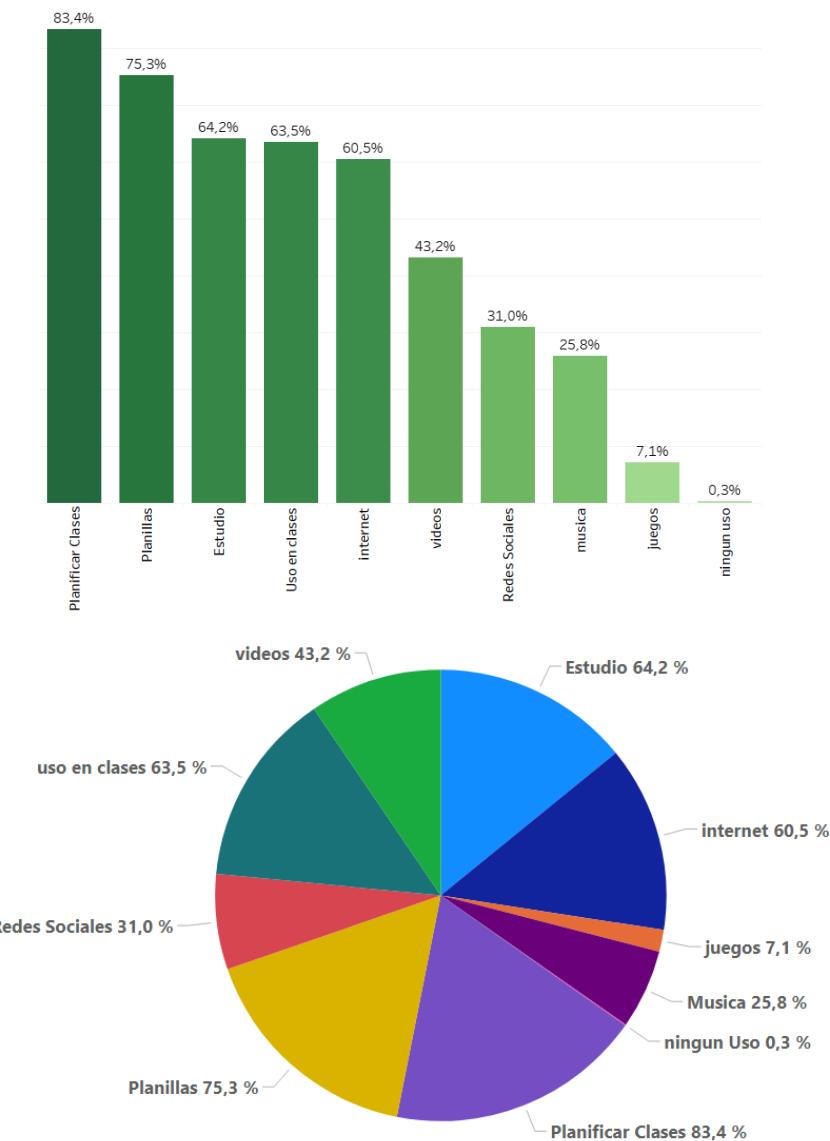
La figura 4-4 se muestra que aparte del uso común del celular para llamadas, los maestros en un 83,1% afirman que usan el celular para navegar en internet, seguido de redes sociales en un 70,1% y solo como juegos en 18,3%.



**Figura 4-4: Uso del celular.**

Fuente: Elaboración propia Tableau, 2023.

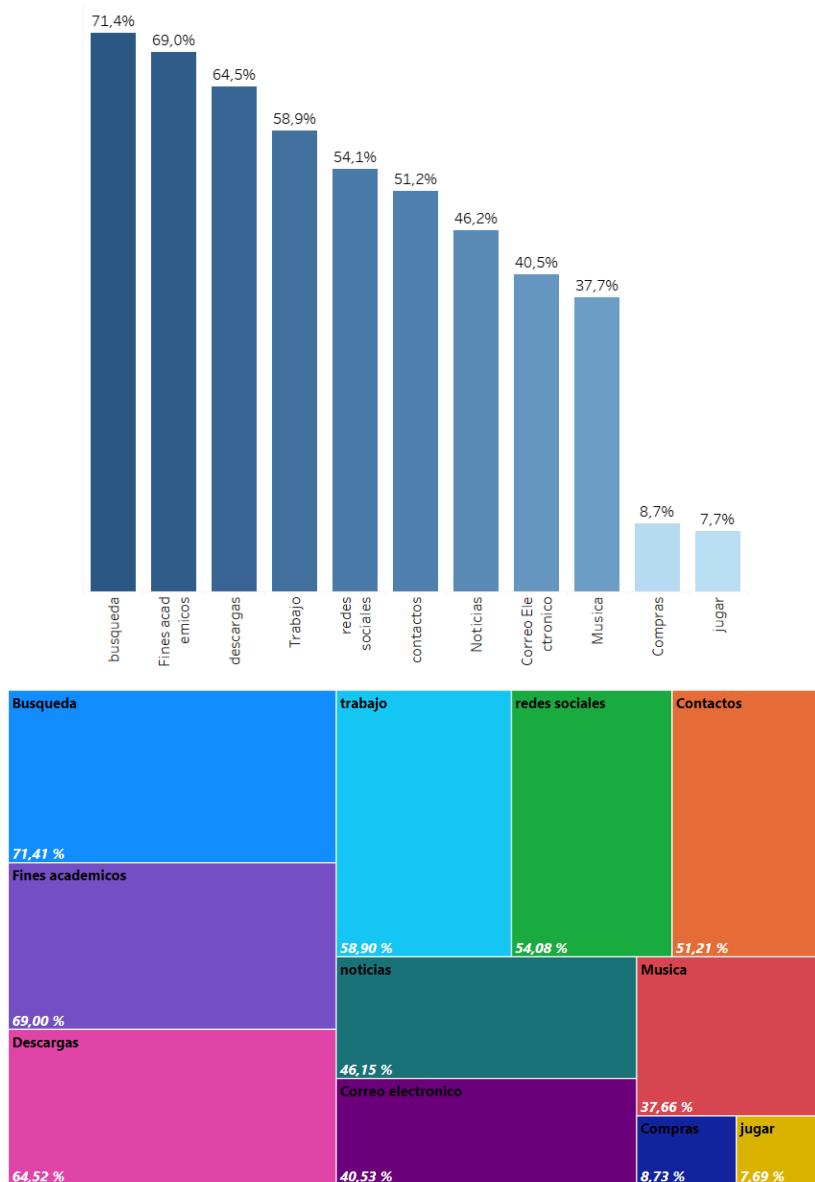
La figura 4-5, muestra gráficamente que los maestros usan la computadora en mejoramiento de la educación, con un 83,4% de los maestros afirmando que usan la computadora para planificar sus clases, y un 63,5% de maestros en uso de computadoras en clases y solo un 7,1% de los maestros afirman usarlo para juegos o diversión.



**Figura 4-5: Uso de la computadora.**

Fuente: Elaboración propia Tableau, 2023.

En la figura 4-6, se muestra la importancia del internet en los maestros y como la utilizan. Un 71.4% de los maestros afirman usar el internet como medio de búsqueda de información, y un 69% lo utiliza para fines académicos, seguidos por descargas, trabajo, redes sociales, etc, y solo un 7,7% de los maestros usan el internet con fines de diversión como video juegos.



**Figura 4-6: Uso del internet en los profesores.**

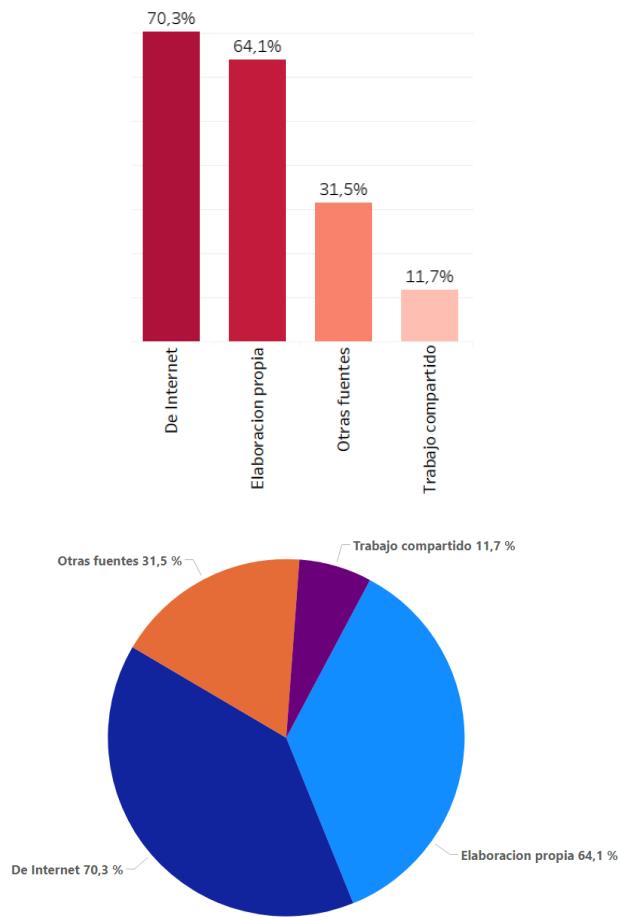
**Fuente:** Elaboración propia Tableau, 2023.

Se puede observar que en su gran mayoría el celular está siendo utilizado para internet y redes sociales, descartando las utilidades más comunes como llamadas, mensajes y demás. No viendo esto como algo negativo sino al contrario que nos indica que los maestros están siempre conectados en internet, para recaudar información más actual y poder añadirlas a sus enseñanzas, y las redes sociales están siendo utilizadas como ayuda para mejorar la educación.

La computadora se ve en su gran mayoría que está siendo utilizada para el ámbito educativo, como llenar planillas, planificar clases, uso en clases y demás, mostrando una utilidad muy buena y aprovechable para la enseñanza.

El internet un servicio primordial hoy en día para la educación, se muestra en los maestros como ayuda para recaudar información, para fines académicos, uso en el trabajo, redes sociales y demás, mostrando como este servicio está siendo de mucha utilidad para los maestros.

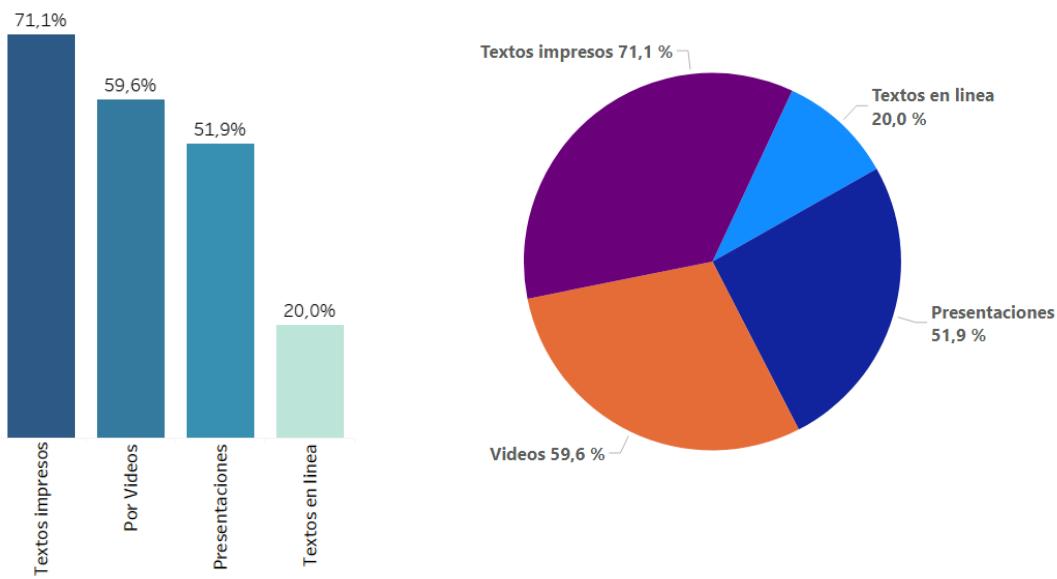
En la figura 4-7, se muestra como el 70,3 % de los maestros evaluados realizan la obtención del material para sus clases del internet, mostrando nuevamente la importancia de contar con internet. El 64,1% de los maestros afirma que ellos mismo realizan sus materiales, y solo el 31,5% de los maestros recaudan su material de otras fuentes y solo el 11,7% lo obtiene de forma compartida con sus colegas.



**Figura 4-7: Preparación del material de los maestros.**

Fuente: Elaboración propia Tableau, 2023.

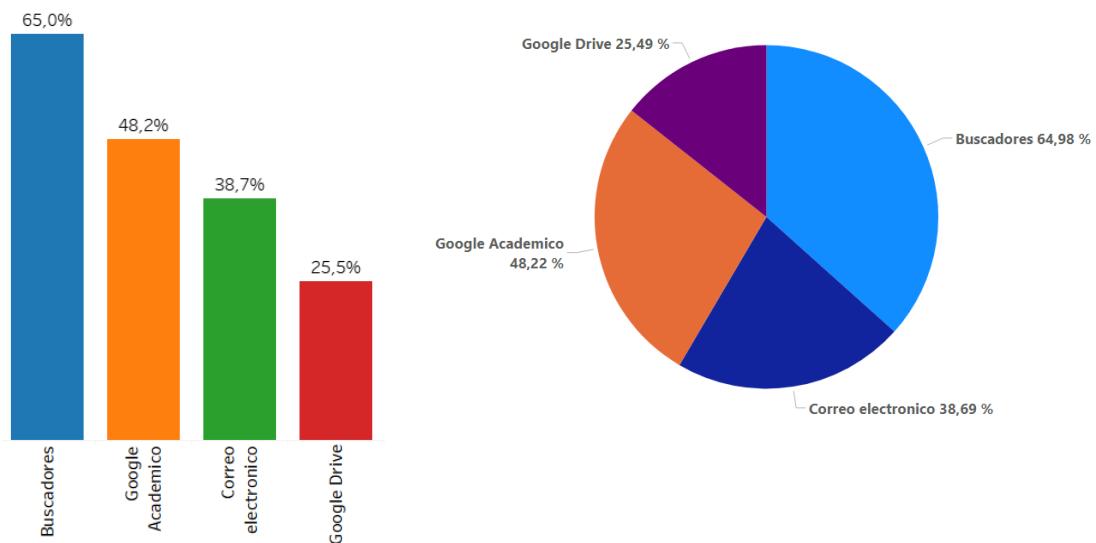
La figura 4-8, muestra que en un 71,1 % de los maestros afirman que sus estudiantes todavía se preparan o estudian de forma clásica que son por medio textos impresos, y el 59,6% de los maestros afirman que sus estudiantes estudian mediante videos informativos y el 51,9% por medio de presentaciones, y solo el 20% afirma que los estudiantes usan textos en línea. Da a entender que los estudiantes todavía necesitan implementar las tecnologías a sus métodos de estudio y aprendizaje.



**Figura 4-8: Resultado de preparación de los estudiantes.**

Fuente: Elaboración propia Tableau, 2023.

En la figura 4-9, se muestra que los maestros solo saben usar el buscador de Google mas no así las otras herramientas como Google académico y Google Drive. Se ve que solo el 25,5% de los maestros saben y usan Google Drive, y solo el 48,2% manejan Google académico. Se ve la necesidad de capacitación a los maestros en conocimiento y uso de herramientas de Google ya que estos son un buen apoyo para la educación de los estudiantes.



**Figura 4-9: Resultado de conocimientos en herramientas de Google.**

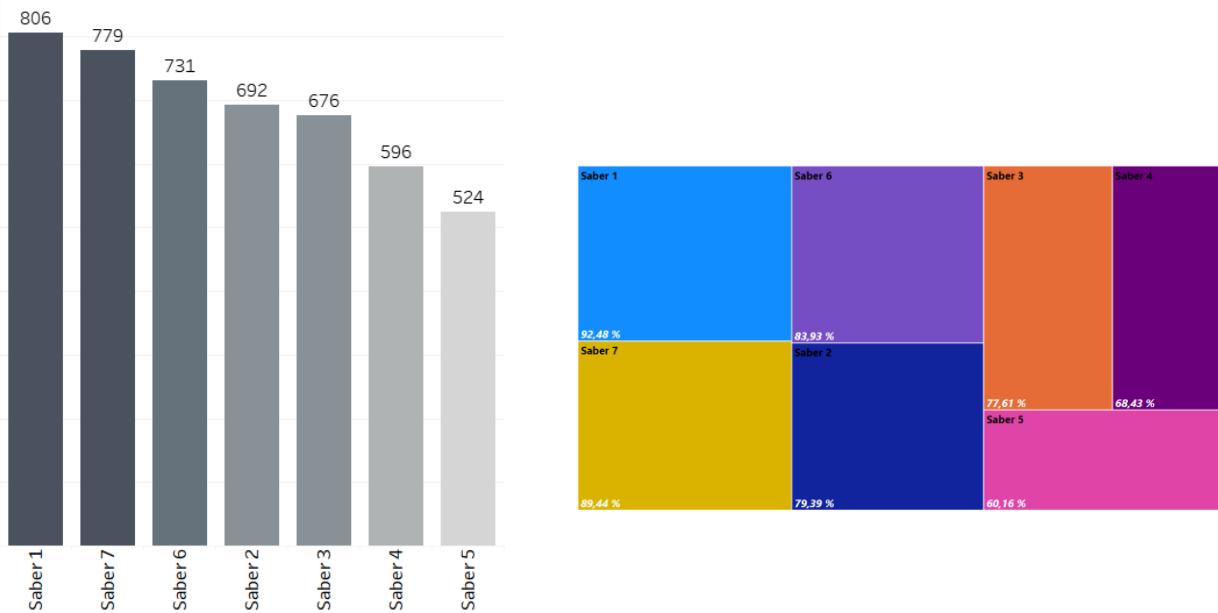
Fuente: Elaboración propia Tableau, 2023.

En la figura 4-10, se muestra que más del 50% de los maestros afirman tener conocimiento de las variables “saber 1”, “saber 2”, “saber 3”, “saber 4”, “saber 5”, “saber 6”, “saber 7”, dando a entender que en su mayoría los maestros de todos los departamentos afirman tener conocimientos básicos de informática tal y como se describe en la tabla 4-1.

**Tabla 4-1: Descripción de las variables saber 1,2,3,4,5,6,7.**

Saberes	Descripción
Saber 1	elaborar, copiar o desplazar texto en un documento y/o buscar, copiar y mover archivos
Saber 2	enviar y abrir correos electrónicos con archivos adjuntos?
Saber 3	conectar o instalar nuevos dispositivos y/o transferir archivos entre la computadora y otro dispositivo
Saber 4	buscar, instalar y configurar programas y/o aplicaciones desde su celular?
Saber 5	buscar, instalar y configurar programas y/o aplicaciones desde su computadora?
Saber 6	conectarse a una red wifi desde una laptop
Saber 7	conectarse a una red wifi desde un teléfono móvil

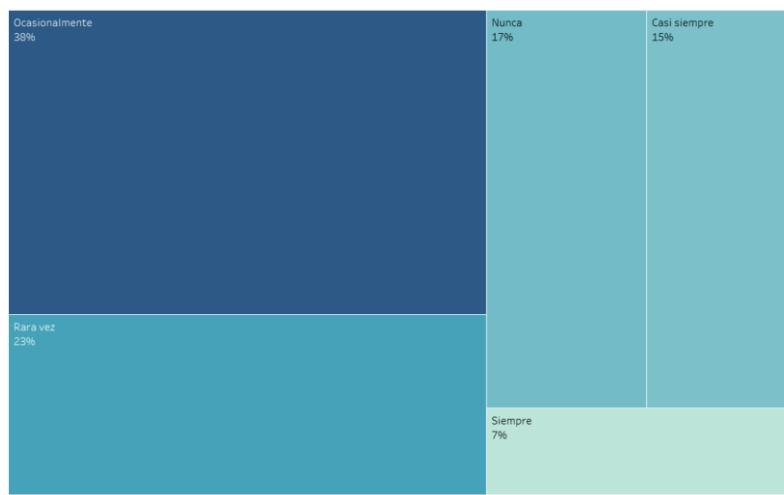
Fuente: Elaboración propia Word, 2023.



**Figura 4-10: Resultado de conocimientos básicos en informática.**

Fuente: Elaboración propia Tableau, 2023.

En la figura 4-11, muestra que el 38% de los maestros afirman implementar las TIC de manera ocasional “ocasionalmente”, seguidos con 23% que afirma la implementación de manera esporádica “Rara vez”, y el 17% afirma que no realiza la implementación “Nunca”, y el 15% lo hace de manera recurrente “casi siempre”, pero solo el 7% afirma que lo hace de manera continua “Siempre”.

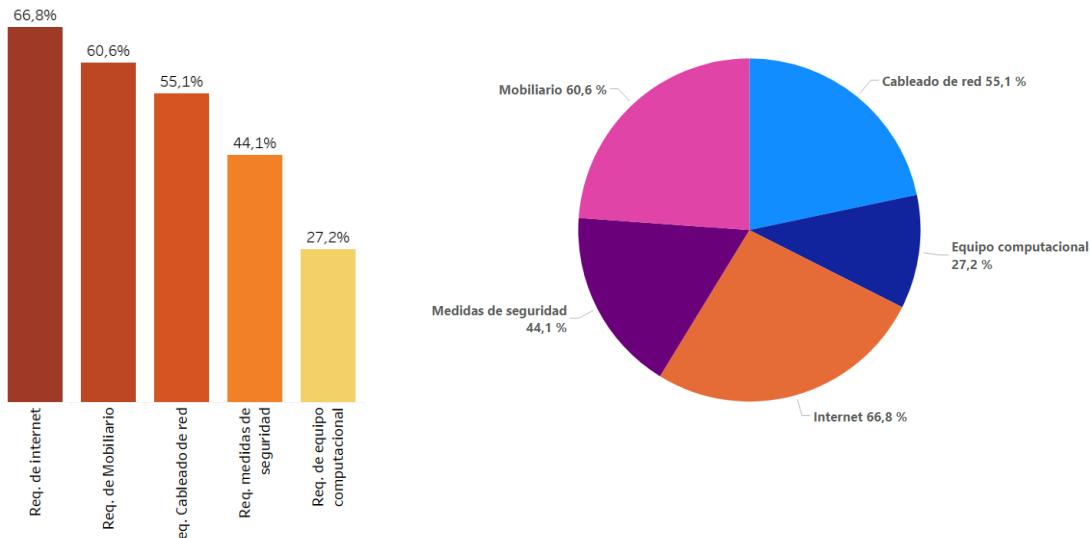


**Figura 4-11: Resultado de implementación de TIC en los colegios.**

Fuente: Elaboración propia Tableau, 2023.

En la figura 4-12, se muestra las necesidades que las escuelas públicas requieren para poder mejorar la educación. Como primer y gran requerimiento se ve el internet con un 66,8% de los maestros que afirman necesitarlo, dando a entender que hay muchísimas escuelas que no pueden avanzar en sus métodos de enseñanza debido a este factor tan importante que es el internet.

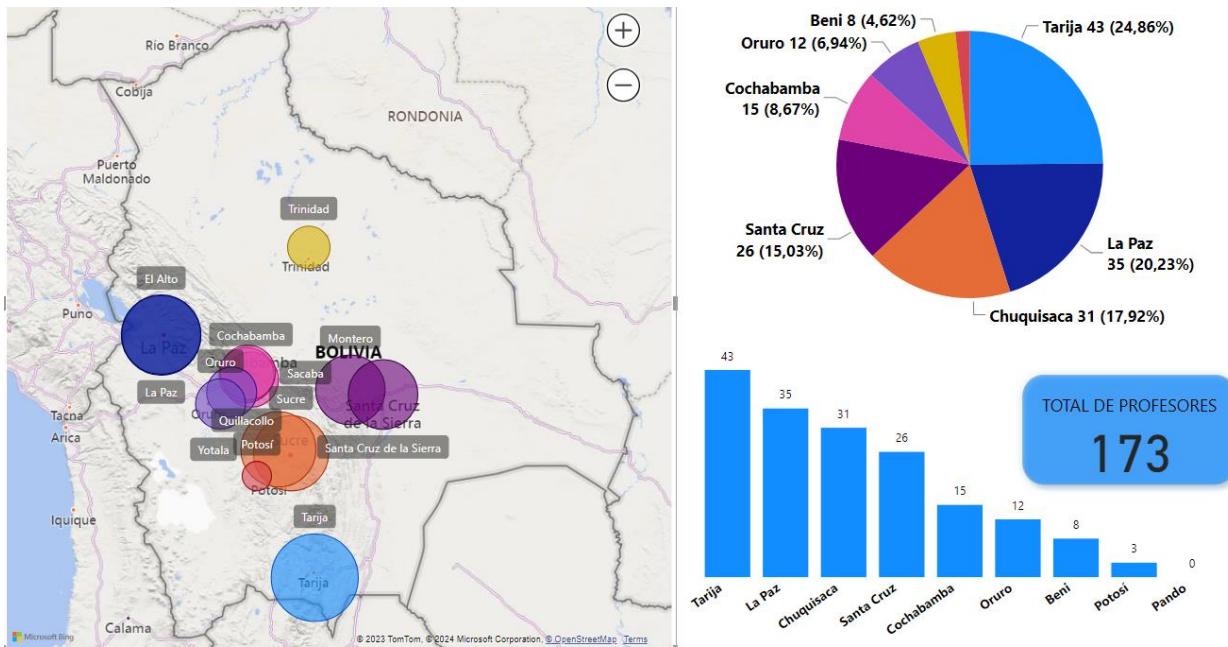
También se observa que más del 50% de los maestros afirman tener la necesidad de poder contar con requerimientos de muebles para su escuela y el 55,1% afirman que requieren cableado de red para todas sus alas del colegio.



**Figura 4-12: Resultado de requerimiento de los colegios.**

Fuente: Elaboración propia Tableau, 2023.

Según análisis descriptivo del conjunto de datos en Power Bi, en la figura 4-13 se puede observar el impacto de la inclusión digital que hubo en cada departamento, siendo Tarija el departamento con más profesores y sus escuelas con impacto de tipo alto. El total de profesores con sus respectivas escuelas que calificaron como inclusión digital de tipo alto fueron solamente 173 de 818 registrados. Con 24,86% de 173 Tarija ocupa el primer lugar en ser catalogado con inclusión digital de tipo alto.



**Figura 4-13: Impacto de inclusión digital de tipo Alto en los 9 departamentos de Bolivia.**

Fuente: Elaboración propia Power Bi, 2023.

En la figura 4-14, se puede observar los departamentos con más inclusión digital de tipo bajo, siendo el departamento de La Paz con 54,88% el más frecuente, seguido por Tarija con 12,09%, donde el departamento de Cochabamba se encuentra con 2,95%. Con 645 de 818 profesores y sus escuelas que fueron categorizados como inclusión digital de tipo bajo.

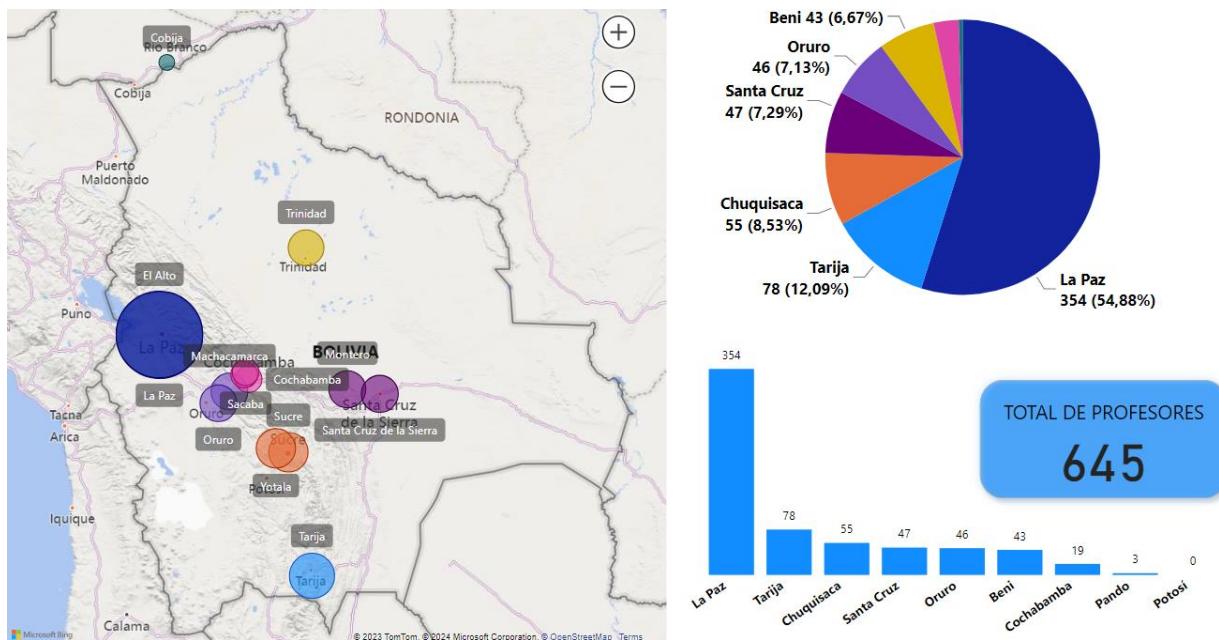


Figura 4-14: Impacto de inclusión digital de tipo Alto en los 9 departamentos de Bolivia.

Fuente: Elaboración propia Power Bi, 2023.

En la figura 4-15, se muestra ya un resumen del tipo de impacto tanto como alto y bajo de los 9 departamentos de Bolivia. Donde Tarija es el departamento con más alto impacto de inclusión digital con 24,86% de 173 y La Paz el más alto en bajo impacto de inclusión digital con 54,88% de 645.

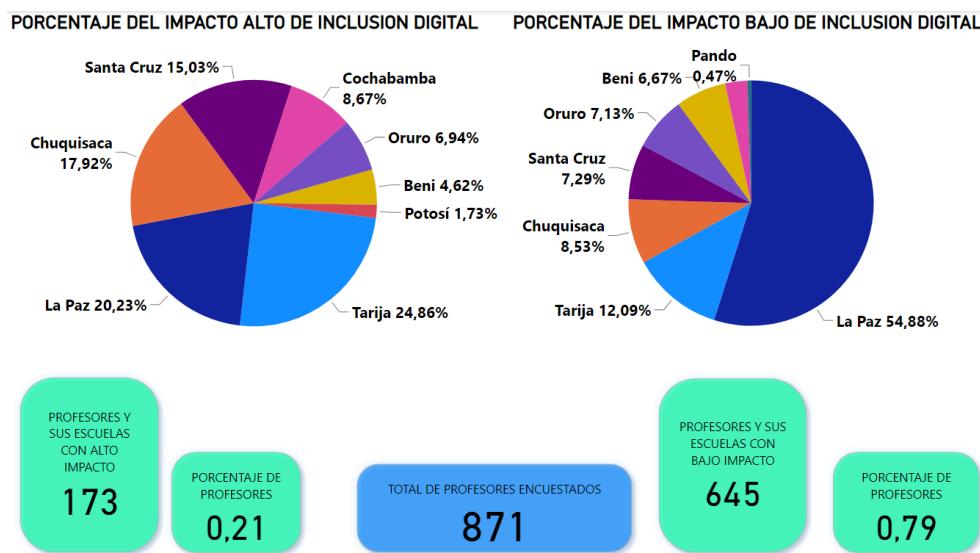


Figura 4-15: Impacto alto y bajo en los 9 departamentos de Bolivia.

Fuente: Elaboración propia Power Bi, 2023.

En la figura 4-16, se muestra según la cantidad de profesores que participaron el porcentaje que hay sobre el impacto de inclusión digital, donde el departamento de Cochabamba llega a ser el departamento donde

el 44% de los profesores participantes tienen un alto impacto en inclusión digital, seguidos por los departamentos de Chuquisaca, Santa Cruz y Tarija con 36%, y el departamento de La Paz solo el 9% de todos los profesores tiene calificación de alto impacto y el 91% son de calificación de bajo impacto. El departamento de Potosí tiene el 100% de calificación de alto impacto al ser solamente 3 profesores y sus escuelas quienes participaron de la encuesta, y el departamento de Pando tiene el 100% de calificación de bajo impacto al ser también solo 3 participantes de la encuesta.

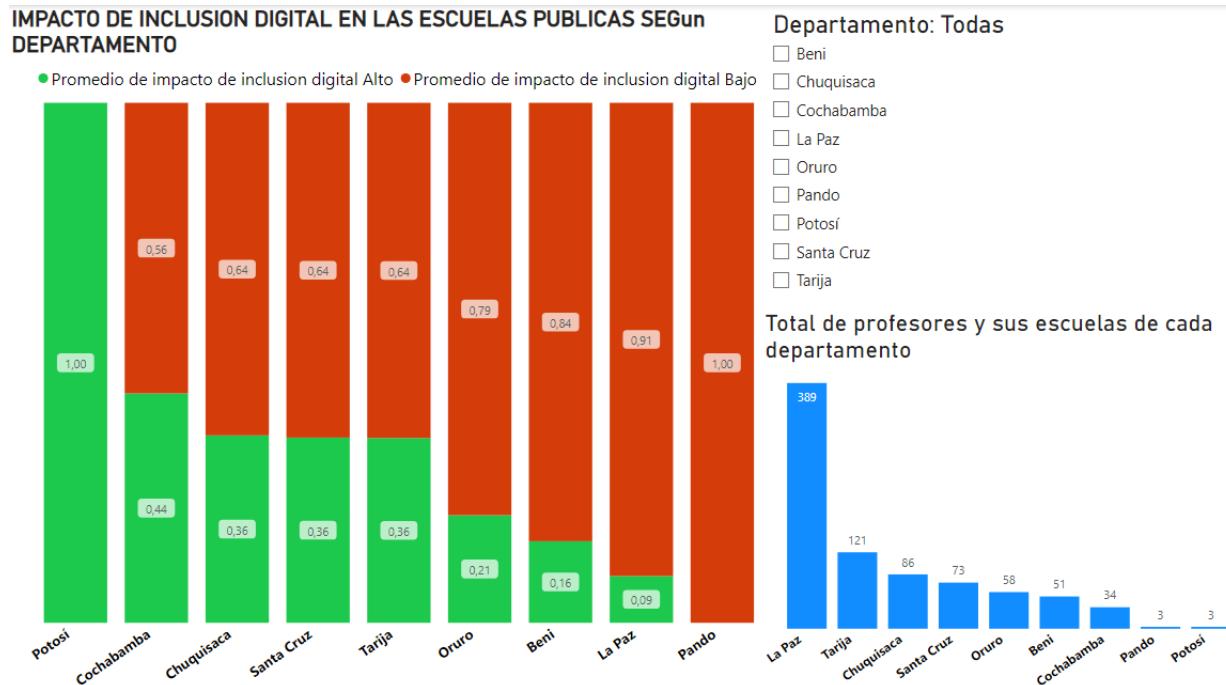


Figura 4-16: Porcentaje de inclusión digital según la cantidad de participación de profesores.

Fuente: Elaboración propia Power Bi, 2023.

## 4.2 Resultados del modelo de clasificación de aprendizaje supervisado de machine learning.

Para determinar un buen modelo de clasificación del impacto de inclusión digital en las escuelas públicas de Bolivia, se hizo el modelado en con varios algoritmos como ya se vio en el capítulo 3. Para poder determinar que algoritmo fue el mejor y el que se debe usar se evaluó las métricas de cada algoritmo.

Se tomaron en cuenta las siguientes métricas:

### 4.2.1. Exactitud (Accuracy).

Fórmula:  $(TP + TN) / (TP + TN + FP + FN)$ . Mide la proporción de predicciones correctas sobre el total de predicciones. Es útil cuando las clases están balanceadas, pero puede ser engañosa en conjuntos de datos desequilibrados.

#### **4.2.2. Precisión Positiva (Precision).**

Fórmula:  $TP / (TP + FP)$ . Indica la proporción de instancias positivas correctamente clasificadas respecto al total de instancias que el modelo ha clasificado como positivas. Es útil cuando el costo de los falsos positivos es alto.

#### **4.2.3. Recall (Sensibilidad o Exhaustividad).**

Fórmula:  $TP / (TP + FN)$ . Mide la proporción de instancias positivas correctamente clasificadas respecto al total de instancias reales positivas. Es útil cuando el costo de los falsos negativos es alto.

#### **4.2.4. F1-Score.**

Fórmula:  $2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$ . Combina precisión y recall en una sola métrica. Es útil cuando hay un desequilibrio entre las clases.

#### **4.2.5. AUC-ROC (Área bajo la curva ROC).**

Representa la capacidad del modelo para discriminar entre clases. Un AUC-ROC más cercano a 1 indica un mejor rendimiento.

#### **4.2.6. Curva ROC (Receiver Operating Characteristic).**

La tabla 4-2, se muestra los valores de las métricas de evaluación de los algoritmos usados para los modelos, los cuales son de mucha importancia al momento de elegir el modelo más adecuado para la clasificación del impacto de inclusión digital.

**Tabla 4-2: Métricas de algoritmos.**

METRICAS DE LOS ALGORITMOS					
ALGORITMO	Accuracy	Precision	Recall	F1-score	AUC-ROC
Logistic Regression	0,971	0,95	1,0	0,974	0,997
K Neighbors Classifier	0,885	0,85	0,94	0,900	
Máquinas de Soporte Vectorial (SVM)	0,942	0,904	1,0	0,950	0,989
Decision Tree Classifier	0,957	0,972	0,94	0,959	0,970
Gaussian NB	0,971	0,95	1,0	0,974	0,981
Redes Neuronales ANN	0,985	0,974	1,0	0,987	0,997

**Fuente: Elaboración propia, 2023**

En la figura 4-17, se puede observar los resultados de la matriz de confusión de los algoritmos árboles de decisión, máquinas de vectores de soporte (SVM), clasificación de Naive Bayes, K vecinos más cercanos (KNN), regresión logística y redes neuronales utilizados para el modelado. Dando una perspectiva más clara y de ayuda para elegir el algoritmo correcto.

Regresión Logística	Positivo (1)	Negativo (0)
Positivo (1)	TP = 30	FP = 2
Negativo (0)	FN = 0	TN = 38

Naive Bayes NB	Positivo (1)	Negativo (0)
Positivo (1)	TP = 30	FP = 2
Negativo (0)	FN = 0	TN = T38

K vecinos cercanos	Positivo (1)	Negativo (0)
Positivo (1)	TP = 26	FP = 6
Negativo (0)	FN = 2	TN = 36

Máquinas de Soporte Vectorial (SVM)	Positivo (1)	Negativo (0)
Positivo (1)	TP = 28	FP = 4
Negativo (0)	FN = 0	TN = 38

Árbol de decisiones	Positivo (1)	Negativo (0)
Positivo (1)	TP = 31	FP = 1
Negativo (0)	FN = 2	TN = 36

Redes Neuronales ANN	Positivo (1)	Negativo (0)
Positivo (1)	TP = 31	FP = 1
Negativo (0)	FN = 0	TN = 38

Figura 4-17: Matriz de confusión de algoritmos de clasificación en aprendizaje supervisado.

Fuente: Elaboración propia, 2023.

#### 4.3 Elección del modelo para clasificación del impacto del programa de inclusión digital.

Según los valores obtenidos de las métricas de los algoritmos como árboles de decisión, máquinas de vectores de soporte (SVM), clasificación de Naive Bayes, K vecinos más cercanos (KNN), regresión logística y redes neuronales vistos en la tabla 4-2; las métricas de Accuracy, Precision, Recall, F1-score, del modelo de clasificación con algoritmo de redes neuronales ANN son los valores más precisos, indicando que es el mejor modelo para la clasificación del impacto de inclusión digital que los demás modelos con los otros algoritmos.

La matriz de confusión del modelo de clasificación con algoritmo de redes neuronales muestra según los valores de falsos positivos como logra equivocarse solo una vez, mientras que en los valores de falsos negativos no se equivoca ni una sola vez, vistos en la figura 4-17, y da a entender que el modelo con

algoritmos de redes neuronales es el más adecuado para la clasificación del impacto de inclusión digital en los departamentos de Bolivia.

En la tabla 4-3, podemos observar la elección de nuestro algoritmo de aprendizaje supervisado de machine learning según el valor de sus métricas de evaluación y de evaluación AUC-ROC.

**Tabla 4-3: Métricas del algoritmo Redes Neuronales (ANN).**

ALGORITMO	METRICAS DE LOS ALGORITMOS				
	Accuracy	Precision	Recall	F1-score	AUC-ROC
Redes Neuronales (ANN)	0,985	0,974	1,0	0,987	0,997

**Fuente:** Elaboración propia, 2023.

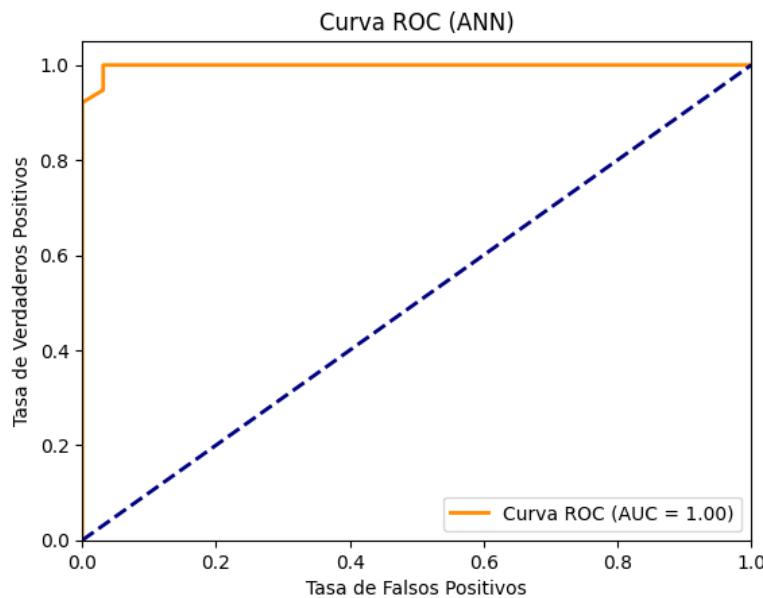
En la tabla 4-4, se observa la matriz de confusión del algoritmo Decision Tree con solo una predicción errónea.

**Tabla 4-4: Matriz de confusión del algoritmo Redes Neuronales (ANN)**

Redes Neuronales ANN	Positivo (1)	Negativo (0)
	Positivo (1)	Negativo (0)
Positivo (1)	TP = 31	FP = 1
Negativo (0)	FN = 0	TF = 38

**Fuente:** Elaboración propia, 2023.

En la figura 4-18, se observa la curva ROC de nuestro modelo de clasificación “Redes Neuronales ANN”.



**Figura 4-18: Curva ROC (Redes Neuronales ANN).**

**Fuente:** Elaboración propia Google Colab, 2023.

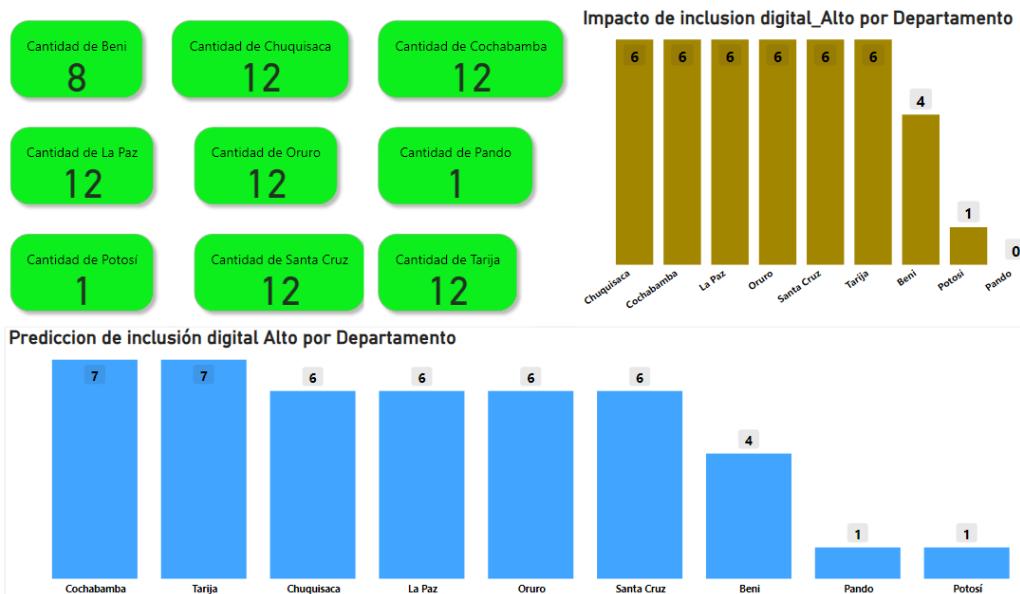
En la figura 4-19, muestra los valores reales de la variable “impacto de inclusión digital” y los valores predichos por nuestro modelo de clasificación de impacto de inclusión digital, mostrando así que el modelo funciona correctamente, con solo un error de clasificación en el registro 780.

	Real	Predicciones	
638	1	1	Calendario
754	1	1	Información
271	0	0	
140	0	0	
456	0	0	
487	1	1	
247	1	1	
413	1	1	
348	0	0	

**Figura 4-19: Valores predichos del modelo (Redes Neuronales ANN).**

Fuente: Elaboración propia Google Colab, 2023.

En la figura 4-20, se muestran los resultados del modelo de clasificación del impacto de inclusión digital con un nuevo conjunto de datos, donde se observa que el modelo se logra equivocar solamente en uno, como en el departamento de Cochabamba 6 de 12 profesores y sus escuelas tienen impacto alto, pero el modelo entrenado clasificó a 7 de 12 como impacto alto equivocándose solamente en uno; Tarija de igual forma clasificó 7 de 12 mostrando un error de 1 y Potosí tuvo un error 1 de 1 ya que solo se encontró un profesor y su escuela. En lo demás el modelo logró clasificar de manera correcta.



**Figura 4-20: Resultados de predicción del nuevo conjunto de datos que se usó como prueba del modelo.**

Fuente: Elaboración propia Power Bi, 2023.

#### 4.4 Discusión de resultados de análisis del conjunto de datos.

Según el estudio realizado por Sulma Farfán Sossa, Antonio Medina Rivilla y María Luz Cacheiro González, para determinar la inclusión digital en la educación de la ciudad de Tarija (Sulma Farfán Sossa, 2015), llegaron a determinar que el 28% de los estudiantes acceden a Internet desde el establecimiento educativo. Asimismo, su estudio ha evidenciado que el espacio más utilizado por los estudiantes para acceder a internet es mediante los café-Internet en un 46%, y solo acceden a internet desde casa en un 25% tal y como se observa en la figura 4-21.

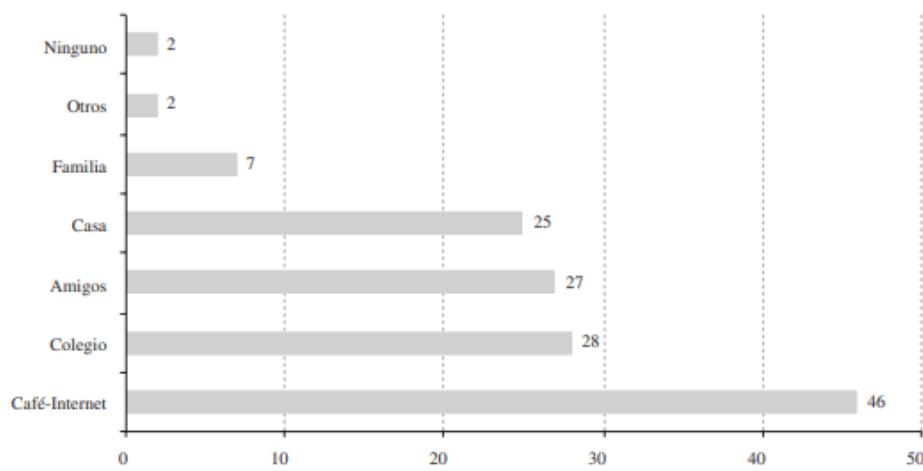


Figura 4-21: Espacios utilizados por los estudiantes para acceder a internet.

Fuente: “Digital inclusion in education in Tarija, Plurinational State of Bolivia”

En el estudio de este proyecto se obtuvo como resultado de la variable “lugares de acceso a internet con más frecuencia” que el 83 % de los maestros acceden a internet desde sus hogares y solo un 8% accede a internet desde la unidad educativa, tal como se observa en la figura 4-22.

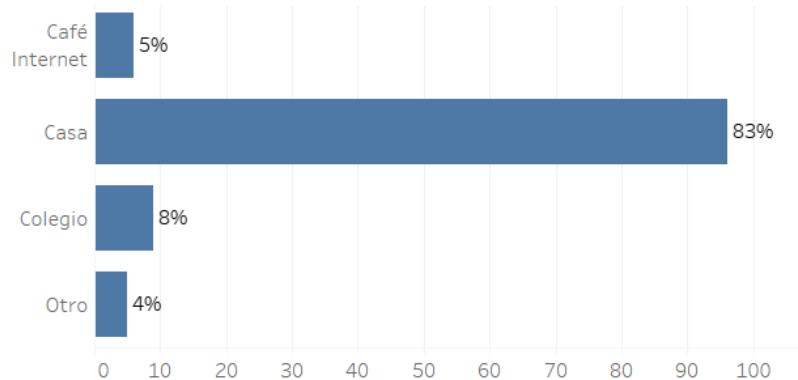
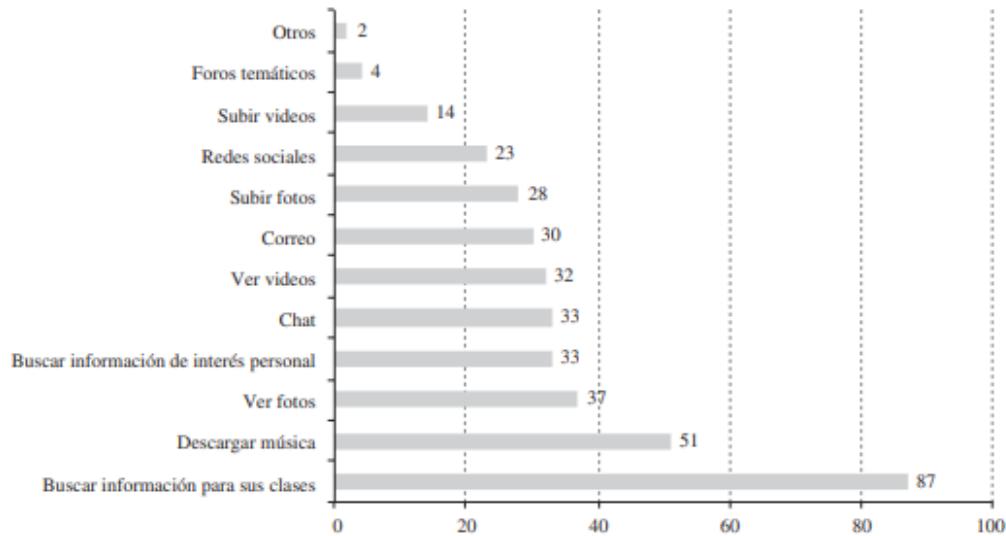


Figura 4-22: Lugares de acceso a internet con más frecuencia.

Fuente: Elaboración propia Tableau, 2023.

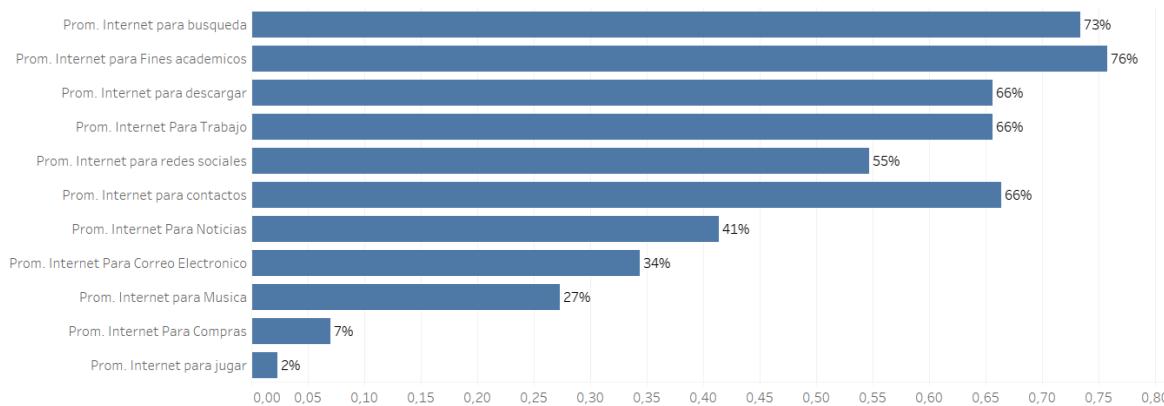
También en el estudio realizado por Sulma Farfán Sossa, Antonio Medina Rivilla y María Luz Cacheiro González (Sulma Farfán Sossa, 2015), muestra como el 87% de los estudiantes manifiestan usar Internet con fines educativos, específicamente en la búsqueda de información. Este porcentaje señala la importancia que tiene el internet como fuente de información para los adolescentes, como se ve en la figura 4-23.



**Figura 4-23: Estudiantes y usos de Internet.**

Fuente: “Digital inclusion in education in Tarija, Plurinational State of Bolivia” (Sulma Farfán Sossa, 2015)

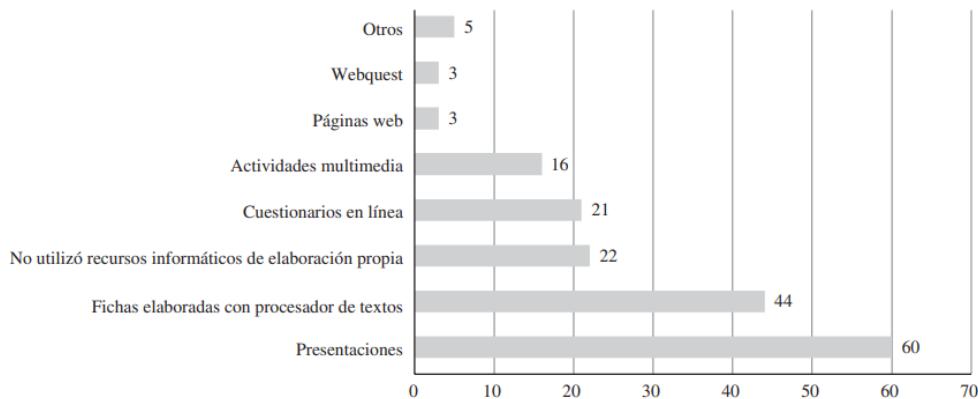
Ante esto el estudio de este proyecto con el análisis de la variable “utilidad del internet” en la figura 4-24, muestra como el 76% de los maestros afirman utilizar el internet con fines académicos, dando a entender que su uso es de mucha importancia al momento de realizar sus enseñanzas, también nos muestra que el 73% lo utilizan para búsqueda de información, y el 66% afirma su utilidad para descargar, trabajo, y contactar a sus amigos. Dejando un 2% para diversión mediante juegos.



**Figura 4-24: Utilidad del internet**

Fuente: Elaboración propia Tableau, 2023.

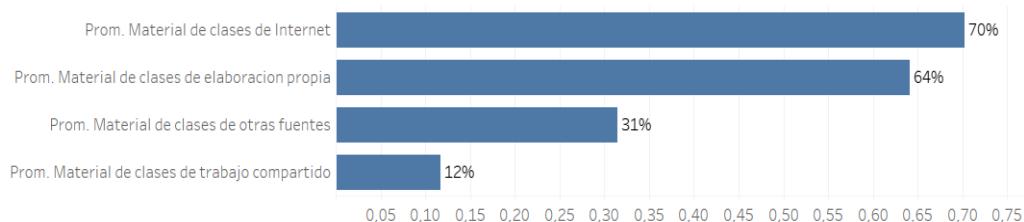
El estudio sobre la inclusión digital en la educación de Tarija por Sulma Farfán Sossa, Antonio Medina Rivilla y María Luz Cacheiro González, muestra en la figura 4-25 como los profesores elaboran diversos recursos para sus clases, basados especialmente en paquetes ofimáticos; no obstante, existe un 22% que declara no usar recursos informáticos de elaboración propia, y un 60% realiza presentaciones como medio de enseñanza para la educación.



**Figura 4-25: Recursos elaborados por los profesores**

**Fuente:** “Digital inclusion in education in Tarija, Plurinational State of Bolivia” (Sulma Farfán Sossa, 2015).

El estudio de este proyecto muestra en la figura 4-26, en el análisis de la variable “Elaboración de material de clases” como el 70% de los maestros afirman que su material o de educación proviene de material de internet, pero también el 64% afirma que su material es elaboración propia y solo el 12% comparte material de educación entre compañeros.



**Figura 4-26: Elaboración del material de clases.**

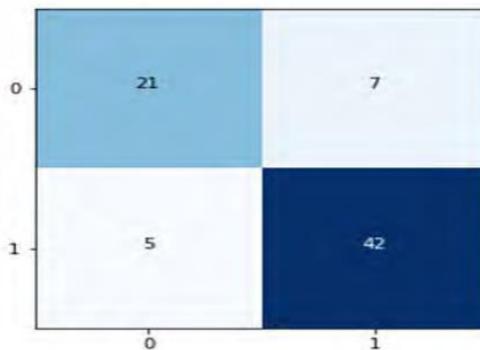
**Fuente:** Elaboración propia Tableau.

Comparando proyectos se puede notar cambios que son debidos a que los análisis del otro autor fueron realizados con información de años menores al 2019; mientras que el estudio de este proyecto maneja información del año 2021, por lo cual es notorio las variaciones en los puntos mostrados del otro autor en sus análisis de datos.

Otro estudio realizado por Juan Francisco Mendoza Bernedo, Fernando Jesús Saldaña Bustamante, Rocío Susana Vivanco Yovera, en su proyecto “Identificación de Obras Urbanas para la Ciudad de Lima

a través del uso de Herramientas basadas en Machine Learning” (Bernedo, 2021) para lo cual usó la red social Twitter para identificar comentarios sobre lugares que necesitan mejorar en la ciudad de Lima.

Para analizar los resultados de las obras públicas relacionadas con parques, consideró aquellos comentarios de Twitter referidos a estas obras. La figura 4-27 muestra el resultado de la matriz de confusión con relación a parques. Asimismo, en la figura 4-28 se aprecian los resultados de la matriz de confusión con referencia a los comentarios y cada uno de sus cuadrantes con sus respectivos comentarios.



**Figura 4-27: Matriz de Confusión Predicción del Modelo en Parques.**

Fuente: “Identificación de Obras Urbanas para la Ciudad de Lima a través del uso de Herramientas basadas en Machine Learning”. (Bernedo, 2021).

Cuadrante	Total	Descripción	Estado de predicción
TN: Verdaderos negativos	21	Numero de comentarios negativos de Twitter referente a parques, que el modelo ha predicho como negativos y que realmente son negativos.	Correcta
FP: Falsos positivos	7	Numero de comentarios positivos de Twitter referente a parques, que el modelo ha predicho como positivos y que realmente son negativos.	Errónea
FN: Falsos negativos	5	Numero de comentarios negativos de Twitter referente a parques, que el modelo ha predicho como negativos y que realmente son positivos.	Errónea
TP: Verdaderos positivos	42	Numero de comentarios positivos de Twitter referente a parques, que el modelo ha predicho como positivos y que realmente son positivos.	Correcta

**Figura 4-28: Descripción de Resultados Predicción del Modelo en Parques.**

Fuente: “Identificación de Obras Urbanas para la Ciudad de Lima a través del uso de Herramientas basadas en Machine Learning” (Bernedo, 2021).

En la evaluación del modelo de clasificación del impacto de inclusión digital en las escuelas públicas realizada en este proyecto, los resultados de la matriz de evaluación del modelo que se muestra en la figura 4-29 indica que sus predicciones son muy buenas, donde solo se puede observar que el modelo se equivocó una sola vez.

Redes Neuronales (ANN)	Positivo (1)	Negativo (0)
Positivo (1)	TP = 31	FP = 1
Negativo (0)	FN = 0	TN = 38

**Figura 4-29: Matriz de evaluación del modelo de clasificación Decision Tree.**

Fuente: Elaboración propia, 2023.

En la figura 4-30, se muestra los valores de las métricas del proyecto “Identificación de Obras Urbanas para la Ciudad de Lima a través del uso de Herramientas basadas en Machine Learning” basados en la evaluación de comentarios sobre parques en la red social Twitter.

Métrica	Resultado
Precision	0,857
Recall	0,893
F1_Score	0,874
Accuracy	0,840

**Figura 4-30: Resultado de Métricas del Modelo de Predicción en Parques.**

Fuente: “Identificación de Obras Urbanas para la Ciudad de Lima a través del uso de Herramientas basadas en Machine Learning” (Bernedo, 2021).

Las métricas de evaluación del modelo de clasificación del impacto de inclusión digital en las escuelas públicas, realizadas en este proyecto como se muestran en la tabla 4-5, nos indica que el modelo seleccionado es el correcto.

**Tabla 4-5: Tabla de las métricas del modelo de clasificación con algoritmo Redes Neuronales (ANN).**

ALGORITMO	METRICAS DE LOS ALGORITMOS				
	Accuracy	Precision	Recall	F1-score	AUC-ROC
Redes Neuronales (ANN)	0,985	0,974	1,0	0,987	0,997

Fuente: Elaboración propia, 2023.

Como se pudo observar en la comparación de las métricas de evaluación tanto del estudio del presente comparando con “Identificación de Obras Urbanas para la Ciudad de Lima a través del uso de Herramientas basadas en Machine Learning”, se puede evidenciar que el modelo de árbol de decisiones al momento de clasificar es el más adecuado.

## 5. Conclusiones

En este proyecto se propuso realizar un modelo que ayude a clasificar el impacto de la inclusión digital en las escuelas públicas de Bolivia, y a su vez rescatar toda la información relevante y de importancia que tanto maestros como sus unidades educativas pueden proporcionar. Ante esto con todo lo realizado en capítulos anteriores, como ser análisis del conjunto de información, programación, visualizaciones, comparaciones y creación de modelo se llegó a las conclusiones.

La recopilación de información de tipo secundaria de los profesores y sus escuelas fue de mucha importancia para la creación del modelo, llegando a recaudarse 871 registros, de los cuales 421 (48,33%) registros son del departamento de La Paz, 128 (14,69%) registros del departamento de Tarija, 89 (10,21%) registros del departamento de Chuquisaca, 79 (9,07%) registros del departamento de Santa Cruz, 59 (6,77%) registros del departamento de Oruro, 53 (6,08%) registros del departamento de Beni, 35 (4,02%) registros del departamento de Cochabamba, 4 (0,05%) registros del departamento de Pando, 3 (0,03%) registros del departamento de Potosí.

Al realizar un AED se encontró relación entre las variables edad del profesor y años de experiencia donde la relación toma el valor de 0,84; indicando que mientras un profesor tenga más edad, tendrá mayor experiencia.

Una de las características sobresalientes de los profesores es que, en un intervalo de 30 a 60 años de edad, los maestros de las escuelas públicas demuestran contar con todo el conocimiento en redes sociales, informática y TIC para poder implementar la inclusión digital en su educación, no siendo así excusa la edad para poder estar con la tecnología a la mano.

Otra característica sobresaliente es que el 73,4% de los maestros evaluados con nivel académico de Licenciatura se sienten capacitados e implementa TIC en sus métodos de enseñanza y solo 17,2 con nivel de Especialidad y 5,3 % con nivel de maestría.

La red social más utilizada por los maestros como apoyo a la educación es WhatsApp con un 94,4% de maestros de afirman usarlo, seguidos por Facebook con un 80,8% y YouTube 64,4%

Con un 83,1% de maestros que afirman usar el celular para búsqueda de información en internet, llegando a si ser el teléfono celular una de las herramientas primordiales para la educación. Así también el 83,4 afirman usar la computadora para la planificación de clases, convirtiendo la computadora como herramienta indispensable para un maestro al momento de realizar sus enseñanzas.

El servicio de internet con un 69% de maestros que afirma usarlo con fines académicos, llega a ser indispensable para ayudar y mejorar tanto la enseñanza como el aprendizaje en maestros y estudiantes.

Así también el 70,3% de los maestros afirman usar el internet para poder preparar su material de educación.

El 71,1% de los maestros afirman que sus estudiantes todavía se preparan y estudian de libros impresos, pese a que hoy en día se cuente con acceso a internet en casi todos los domicilios.

El 65% de los maestros afirman usar Google solo como medio de búsquedas de información y el 48% menciona conocer Google académico y solo el 25 % afirma conocer Drive, haciendo notar que Google es solo una herramienta de búsqueda, y no así una poderosa herramienta de ayuda para la educación.

El 38% de los maestros afirman implementar las TIC en sus clases de manera ocasionalmente y solo el 7% lo hace de manera siempre, y el 17% de ellos afirman no implementar, dando a entender que hay algo que está impidiendo que los maestros puedan implementar con normalidad todos los días.

El 66,8% de los maestros mencionan tener la necesidad de contar con internet en sus unidades educativas, dando a conocer nuevamente la importancia de este servicio tan importante para la enseñanza y educación.

El modelado, entrenamiento y prueba del modelo de clasificación con algoritmo de redes neuronales ANN del impacto de inclusión digital se realizó de manera correcta presentando como resultados los valores de las métricas de evaluación: Accuracy igual a 0,98; Precision igual a 0,97; Recall igual a 1,0; F1-score igual a 0,98; AUC-ROC igual a 0,997; llegando a ser el más apropiado con respecto a los modelos con algoritmos de árbol de decisiones, máquinas de vectores de soporte (SVM), Naive Bayes, K vecinos más cercanos (K-NN), y regresión logística.

De los registros recaudados para el entrenamiento y prueba del modelo de clasificación, fueron 173 los que se consideraron como de alto impacto donde, el departamento de Tarija con 43 (24,86%) escuelas logró ponerse como el mejor con respecto a la implementación de la inclusión digital, seguidos por el departamento de La Paz con 35 (20,23%), y el departamento de Cochabamba con 15 (8,67%).

Tomando en cuenta solo la cantidad de profesores y sus escuelas que participaron de cada departamento, se puede decir que de los 34 profesores que participaron del departamento de Cochabamba solo el 44% tuvo un aprovechamiento de la capacitación del programa de inclusión digital.

La prueba que se realizó al modelo con un conjunto de datos nuevo, dio un resultado de predicción de solo una falla en los departamentos de Cochabamba, Tarija y Potosí al momento de clasificación como alto impacto, tal como la matriz de confusión indica en su valor de FP= 1, llegando a demostrar nuevamente que el modelo funciona correctamente para la clasificación de alto y bajo impacto de la inclusión digital.

## 6. Recomendaciones

Con todo lo visto en los capítulos anteriores se recomienda para un mejor estudio con respecto a la inclusión digital en las escuelas públicas de Bolivia, tomar en cuenta una misma cantidad de maestros de cada departamento, y mejor aún si se toma en cuenta a todos los maestros de todas las escuelas públicas, ya que los datos con los que se trabajaron en este estudio solamente fue de unos cuantos, el recaudar los datos de todos los maestros, de todas las escuelas de todos los departamentos y ciudades ayudaría a mejorar este proyecto.

Se recomienda también recaudar información de los maestros y escuelas actuales del año, para ver si la inclusión digital va mejorando cada año que pasa, o si no las hay para así poder tomar decisiones con respecto al programa de inclusión digital.

Para la elección correcta del modelo de clasificación con machine learning se recomienda realizar el modelado con varios algoritmos de clasificación y no solo enfocarse en uno, para así mediante sus métricas de evaluación se pueda definir de mejor forma la adecuada para clasificar.

Se recomienda recaudar toda la información posible actualizada sobre inclusión digital, brechas digitales, el estado actual de las escuelas públicas en Bolivia, los conocimientos tecnológicos que los maestros adquieren con el pasar de los años, el mejoramiento de la educación con la implementación de los programas que el gobierno va implementando. Todo esto será de apoyo para reforzar el marco teórico.

También se recomienda recaudar información sobre las calificaciones de los estudiantes antes y después de la integración de TIC, con relación a las materias donde los maestros si implementaron la inclusión digital a sus métodos de educación.

Es recomendable poder contar con un conjunto de datos completo para evitar sesgos y que esto pueda provocar fallas en el modelo de clasificación del aprendizaje supervisado.

## Referencias bibliográficas

- Agetic. (2018). *Estado TIC*. La Paz.
- AGETIC. (09 de diciembre de 2019). Obtenido de AGETIC: <https://www.agetic.gob.bo/agetic-lanza-el-programa-inclusion-digital-2021/>
- AGETIC. (19 de mayo de 2019). Obtenido de AGETIC: <https://www.digital.gob.bo/2019/05/inclusion-digital/>
- AGETIC. (9 de Diciembre de 2019). *www.agetic.gob.bo*. Obtenido de <https://www.agetic.gob.bo/agetic-lanza-el-programa-inclusion-digital-2021/>
- AGETIC. (03 de Junio de 2019). *www.agetic.gob.bo*. Obtenido de <https://www.agetic.gob.bo/convocamos-a-voluntarios-y-voluntarias-para-el-programa-de-inclusion-digital/>
- AGETIC. (17 de Mayo de 2019). *www.digital.gob.bo*. Obtenido de [www.digital.gob.bo: https://www.digital.gob.bo/2019/05/se-presenta-el-proceso-de-construccion-de-la-agendadigital-de-bolivia/](https://www.digital.gob.bo/2019/05/se-presenta-el-proceso-de-construccion-de-la-agendadigital-de-bolivia/)
- Arese, A. S. (2016). *INCLUSIÓN DIGITAL EN AMÉRICA LATINA Y EL CARIBE*.
- AYALA, C. F. (2017). *APLICANDO INTELIGENCIA DE NEGOCIOS DE AUTOSERVICIO, UTILIZANDO POWER BI, PARA LA TOMA DE DECISIONES DENTRO DE UNA PYME EN LA REGIÓN DE TACNA*. Tacna-Peru.
- Bernedo, J. F. (2021). *Identificación de Obras Urbanas para la Ciudad de Lima a través del uso de herramientas basadas en Machine learning*. Lima.
- Borja-Robalino, R. (2020). *Estandarización de métricas de rendimiento para clasificadores de machine y deep learning*. Barcelona: RISTI.
- BRECHA CERO. (1 de Agosto de 2019). Obtenido de BRECHA CERO: <https://brechacero.com/bolivia-convoca-a-docentes-y-jovenes-a-participar-en-el-programa-de-inclusion-digital-en-escuelas-publicas/>
- Bustillos, J. C. (2016). *Infraestructura de telecomunicaciones y TIC en Bolivia*. La Paz.
- Carles Sigalés, J. M. (2009). *La integración de internet en la educación escolar española*. Barcelona: Ariel.
- CEPAL. (2020). *América Latina y el Caribe ante la pandemia del COVID-19*.
- Cortina, V. G. (2015). *APLICACIÓN DE LA METODOLOGÍA CRISP-DM A UN PROYECTO DE MINERÍA*. Madrid.
- Elosua, P. (2021). *Impacto de la TIC en el entorno evaluativo. Innovaciones al servicio de la mejora continua*. San Sebastian .
- Esnaola, L. (s.f.). *Uso de herramientas tecnológicas aplicadas a la educación*. Buenos Aires, Argentina.
- Francisco Javier Lena Acebo, M. E. (2021). *Avances en Educación, TIC e innovación: Aportaciones para la mejora empresarial y social*. Madrid.

- Google. (s.f.). <https://colab.research.google.com/?hl=es>. Obtenido de  
<https://colab.research.google.com/?hl=es>.
- Guadalupe Aurora Maldonado Berea, J. G.-R. ( 2019). *El efecto de las TIC y redes sociales en estudiantes universitarios*. Mexico.
- Judith, S. L. (19 de 07 de 2018). *ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS*.
- Lacruz, M. d. (2010). *INDICADORES SOCIALES DE INCLUSIÓN DIGITAL: BRECHA Y CIUDADANÍA DIGITAL*. Zaragoza.
- LLANOS, G. M. (2019). *ANÁLISIS GEOPOLÍTICO DE LAS TIC EN BOLIVIA*.
- Sulma Farfán Sossa, A. M. (2015). *La inclusión digital en la educación*.
- Tableau. (2003). <https://www.tableau.com/es-es/why-tableau/what-is-tableau>. Obtenido de  
<https://www.tableau.com/es-es/why-tableau/what-is-tableau>.

## Anexos

### Anexo 1. Tabla resumen de la encuesta sobre la inclusión digital a maestros de las escuelas públicas de Bolivia.

Sexo	Eda	Departamento	Colegio	Años de experiencia	internet en colegio	Uso de celular en clases
Masculino	32	Santa Cruz	NACIONAL COTOCO	20	Sí	No
Masculino	38	La Paz	PRIMAVERA B	19	No	Sí
Femenino	54	La Paz	DOMINGO FAUSTINO SARMIENTO	20	Sí	Sí
Femenino	54	La Paz	ABRAHAM REYES FE Y ALEGRIA	26	No	Sí
Masculino	34	La Paz	ABRAHAM REYES FE Y ALEGRIA	10	No	No
Masculino	35	La Paz	TUPAC AMARU	5	Sí	No
Masculino	51	La Paz	ABRAHAM REYES FE Y ALEGRIA	21	No	Sí
Femenino	37	La Paz	PUERTO DE MEJILLONES CIUDAD SATELITE	13	Sí	Sí
Femenino	62	Oruro	PANTALEON DALENCE 1	34	Sí	Sí
Femenino	44	Oruro	JUAN MISAEL SARACHO SECUNDARIA	19	No	Sí
Femenino	33	La Paz	ABEL ITURRALDE B	5	Sí	No
Femenino	55	La Paz	PRIMAVERA B	30	No	No
Masculino	38	La Paz	1RO. DE MAYO TARDE	10	No	Sí
Femenino	59	Oruro	PANTALEON DALENCE 1	25	Sí	Sí
Masculino	68	La Paz	1RO. DE MAYO TARDE	25	No	No
Masculino	29	La Paz	CRISTIANO VIDA NUEVA	7	No	Sí
Masculino	39	La Paz	1RO. DE MAYO TARDE	8	No	Sí
Masculino	55	La Paz	ALFREDO VARGAS B	5	No	Sí

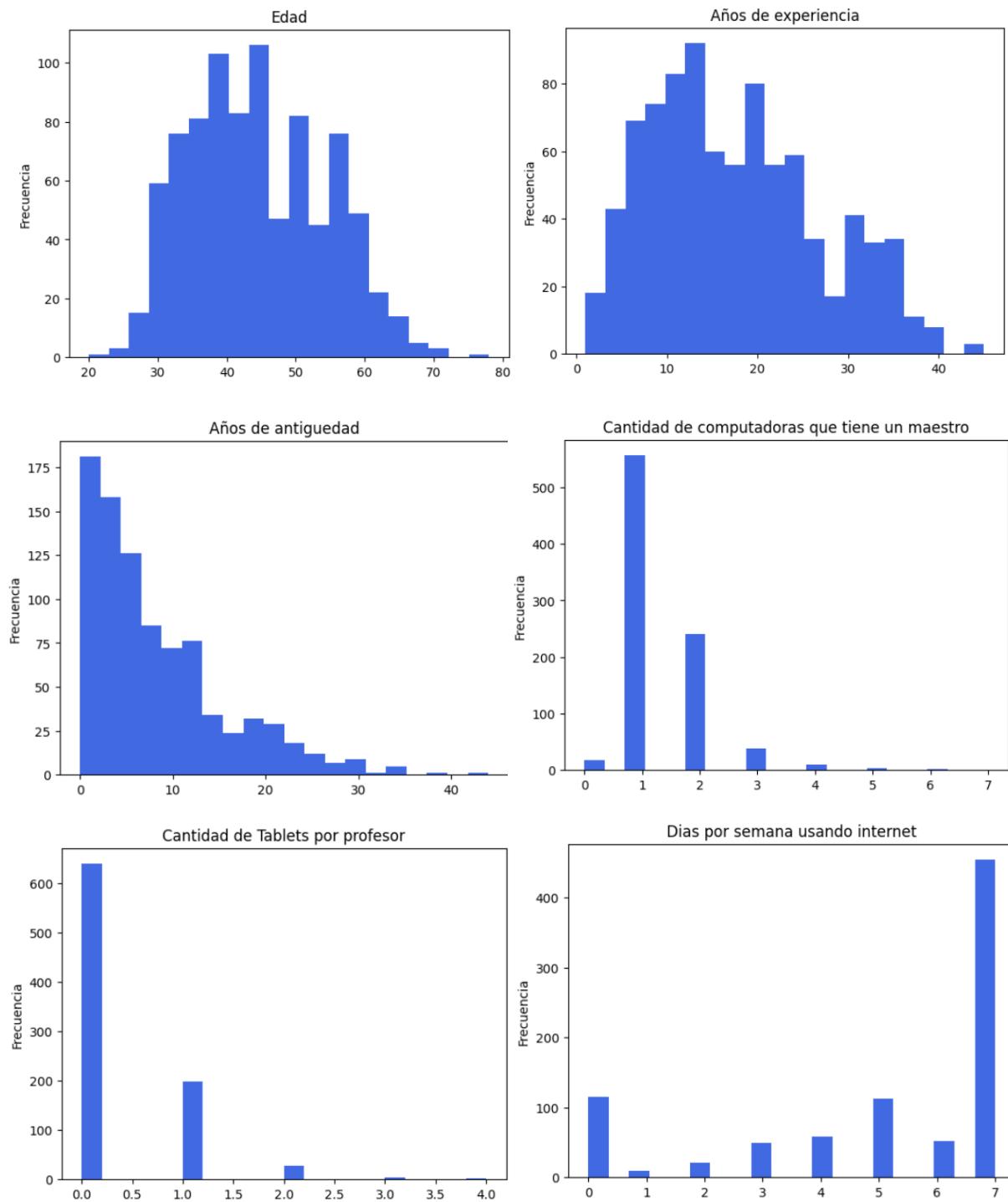
Saber 1	Saber 3	Saber 5	Incorporacion de TIC en aula	Modalidad de Clases	Uso de tecnologias mejoro el aprendizaje
Sí	Dudoso	Dudoso	Casi siempre	Clases frontales	Sí
Sí	Sí	Sí	Siempre	Clases prácticas	Sí
Sí	Sí	Sí	Ocasionalmente	Clases prácticas	Sí
Sí	Sí	Sí	Ocasionalmente	Clases prácticas	Sí
Sí	Sí	Sí	Ocasionalmente	Clases prácticas	Sí
Sí	Sí	Sí	Casi siempre	Talleres	Sí
Sí	Sí	Sí	Rara vez	Clases frontales	Sí
Sí	Dudoso	Dudoso	Ocasionalmente	Clases prácticas	Sí
Dudoso	Dudoso	Dudoso	Casi siempre	Clases prácticas	Sí
Sí	Dudoso	Sí	Ocasionalmente	Clases frontales	Sí
Sí	Sí	Sí	Rara vez	Clases prácticas	Sí
Dudoso	Dudoso	Dudoso	Rara vez	Clases prácticas	No
Sí	Sí	Dudoso	Ocasionalmente	Clases frontales	Sí
Sí	No	No	Rara vez	Clases prácticas	Sí
Sí	Sí	Sí	Casi siempre	Talleres	Sí
Sí	Sí	Sí	Casi siempre	Clases de laboratorios	Sí
Sí	Sí	Sí	Siempre	Clases frontales	Sí
Sí	Dudoso	Sí	Rara vez	Clases prácticas	Sí
Sí	Sí	Dudoso	Nunca	Clases prácticas	Sí
Sí	Sí	Sí	Siempre	Talleres	No

Planilla Completa en: Carpeta llamada DATASET que se encuentra en el CD, o también en la siguiente dirección de Google Drive, en la capeta Dataset en el archivo Excel con el nombre de profesores\_final.

Link:

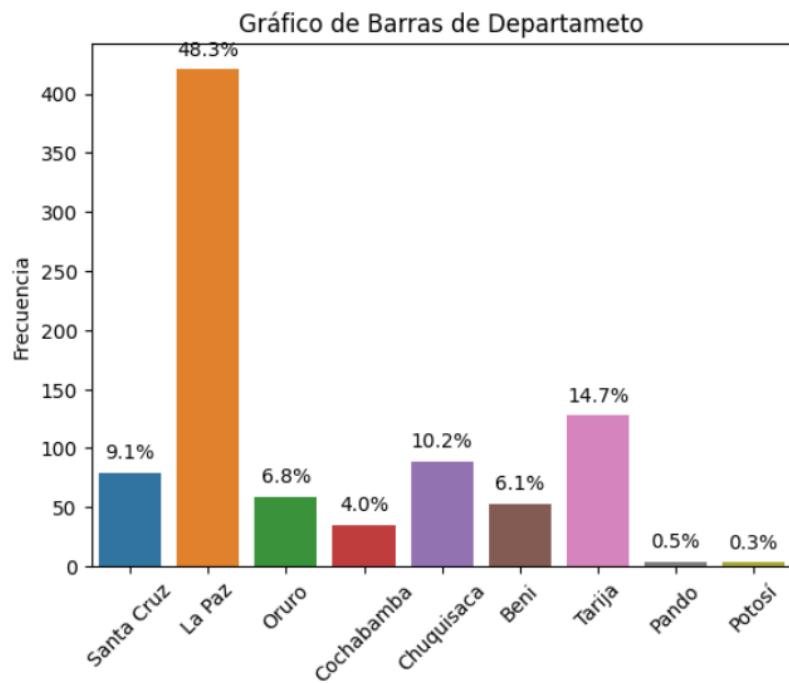
[https://drive.google.com/drive/folders/1jyiRZ4HqAPWbU84RqO6QEeQMdAEVeSR3?usp=drive\\_link](https://drive.google.com/drive/folders/1jyiRZ4HqAPWbU84RqO6QEeQMdAEVeSR3?usp=drive_link)

## Anexo 2. Gráficas de variables cuantitativas.

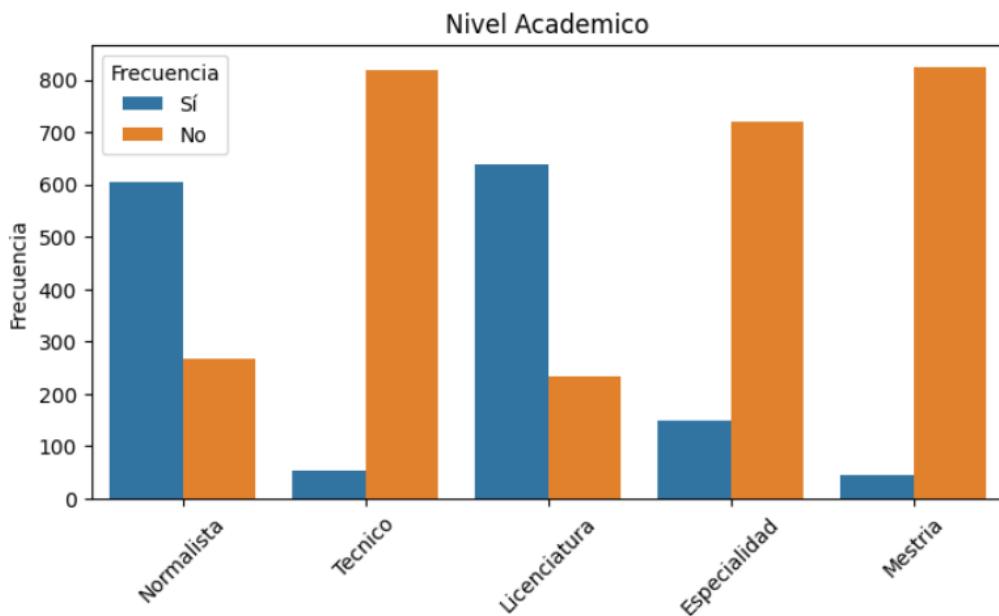


### Anexo 3. Gráficas de variables cualitativas.

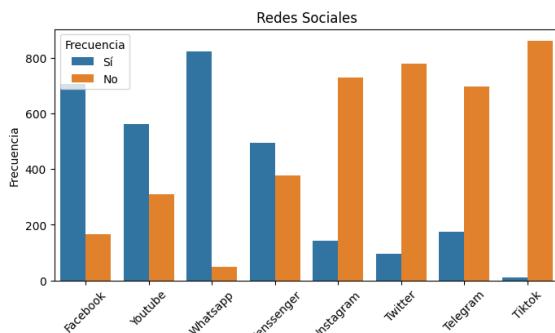
#### Anexo 3.1: Grafico de Barras sobre la frecuencia de los profesores por Departamento.



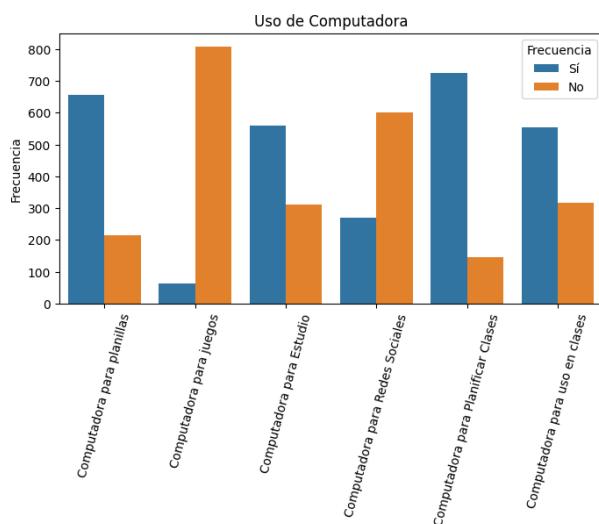
#### Anexo 3.2: Grafico de Barras sobre el nivel académico de los profesores.



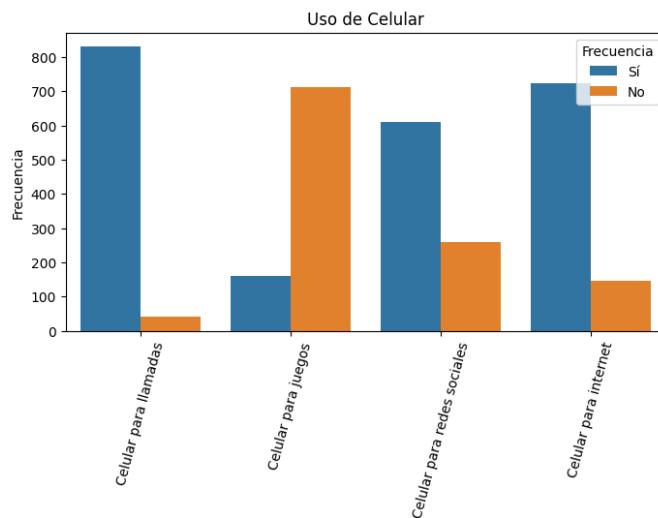
**Anexo 3.3: Gráfica de frecuencia de uso de redes sociales por los profesores.**



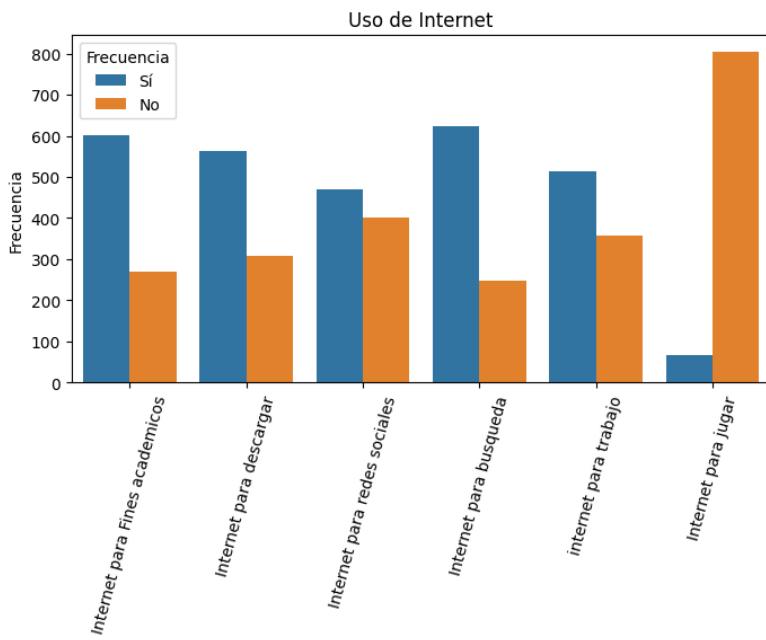
**Anexo 3.4: Grafica de frecuencia de la utilidad de la computadora por los profesores.**



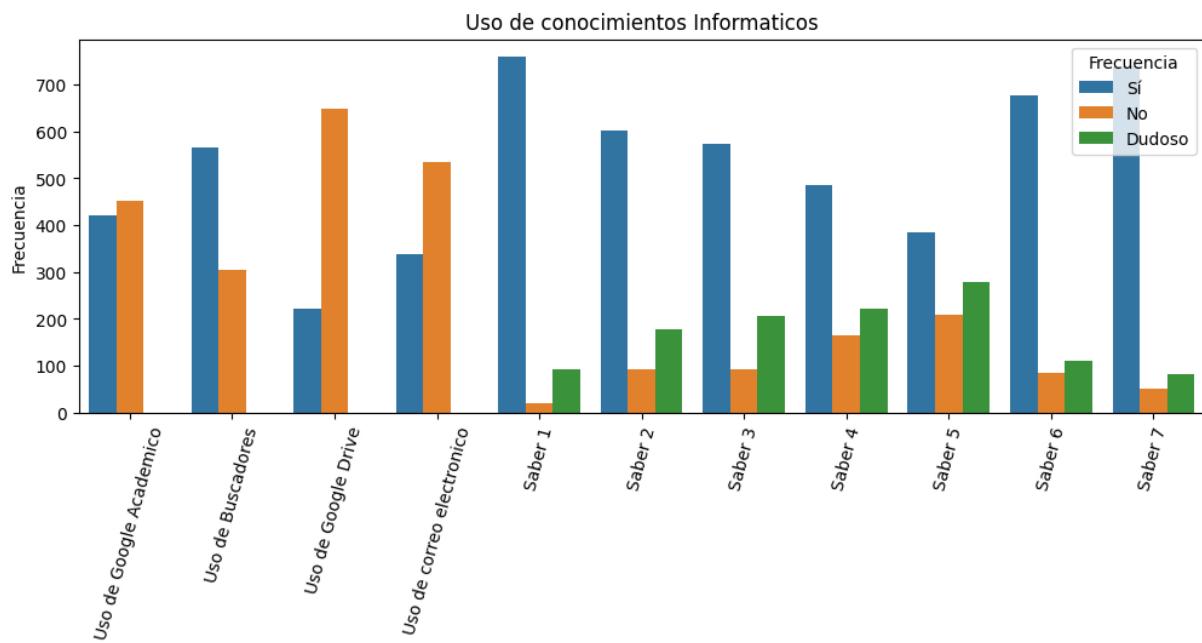
**Anexo 3.5: Grafica de frecuencia de la utilidad del uso de celular por los profesores.**



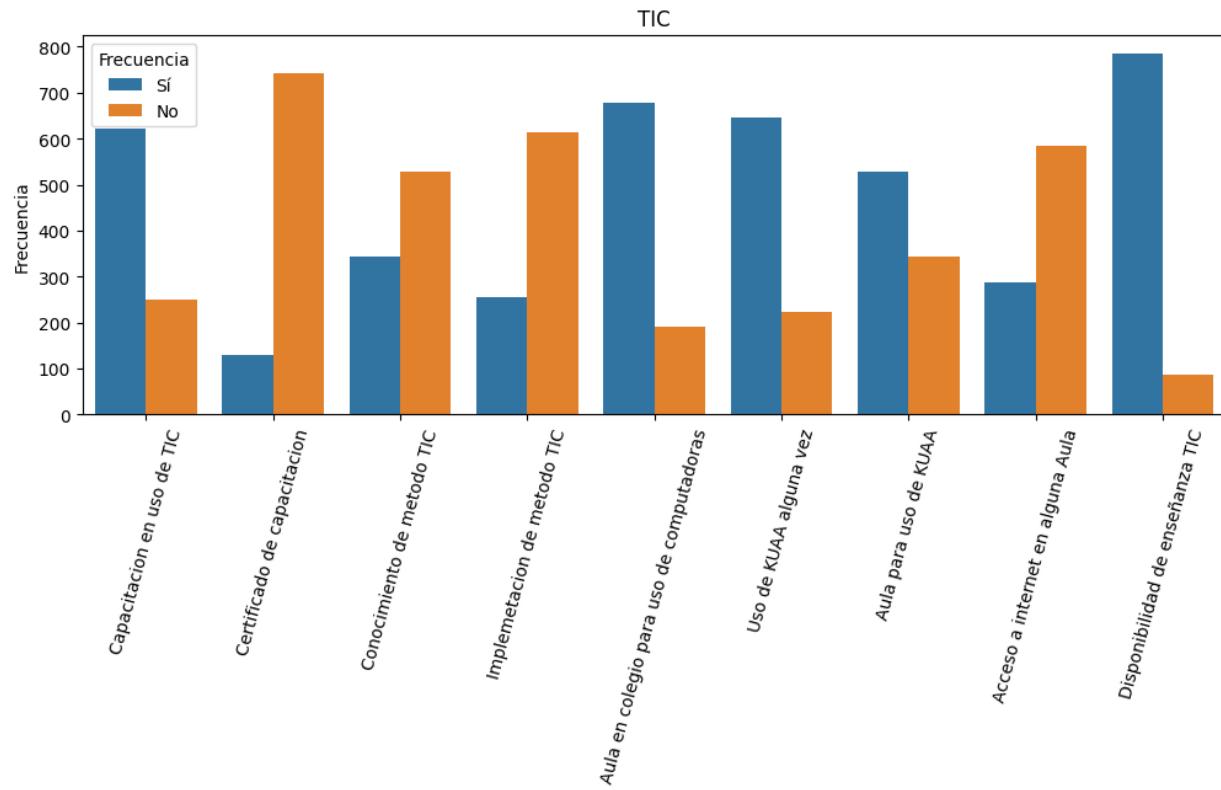
**Anexo 3.6: Grafica de frecuencia de la utilidad del uso de Internet por los profesores.**



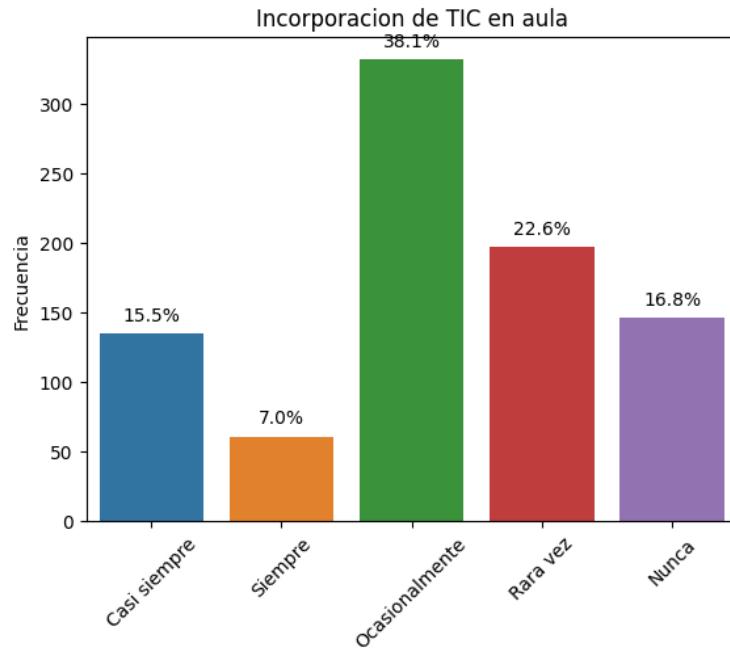
**Anexo 3.7: Grafica de frecuencia del uso de conocimientos informáticos, ofimáticos y herramientas de Google.**



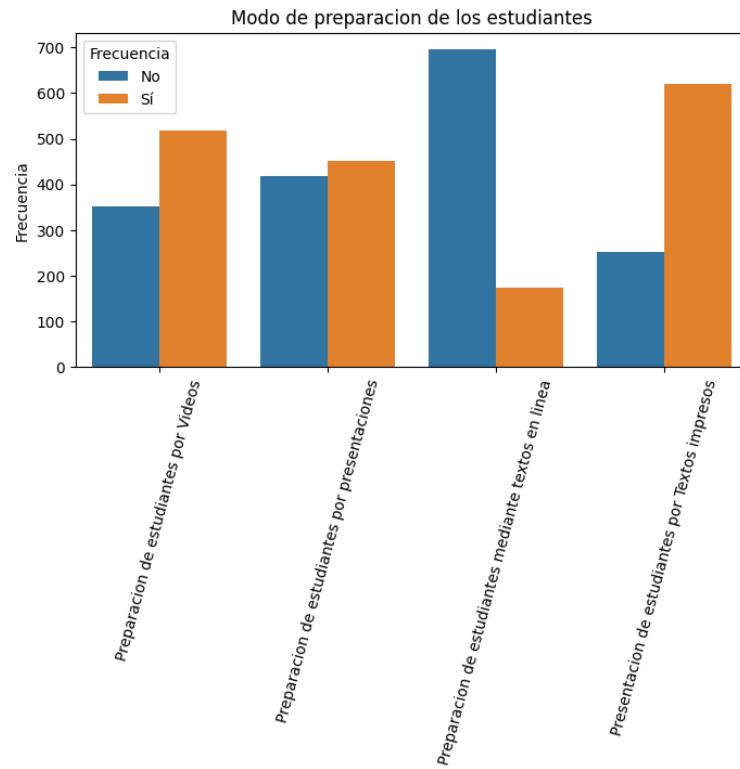
**Anexo 3.8: Grafica de capacitación, conocimiento e implementación de métodos TIC.**



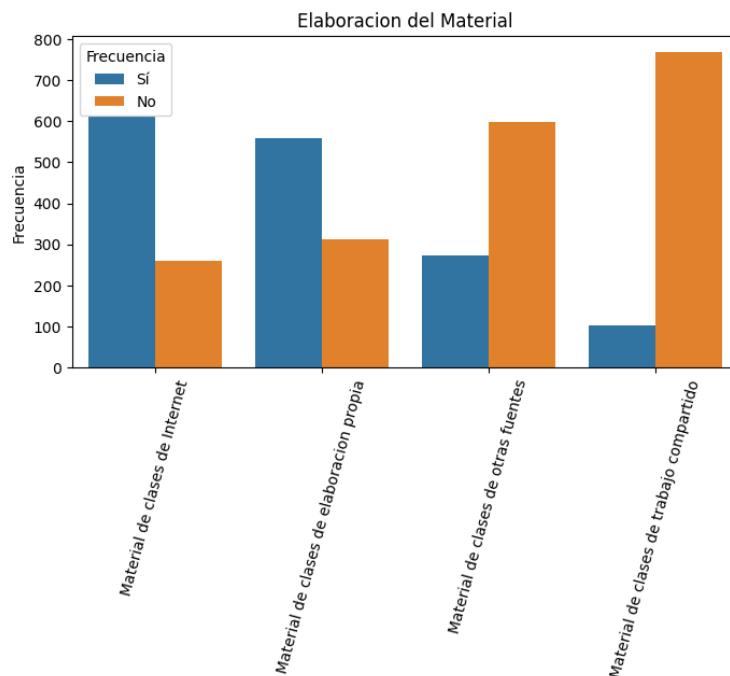
**Anexo 3.9: Grafica de la frecuencia de implementación de TIC en la enseñanza de los profesores.**



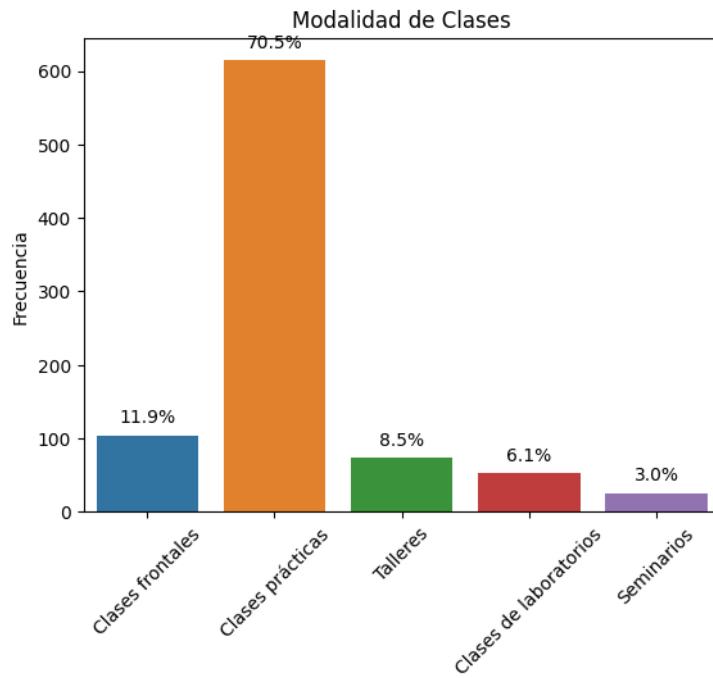
**Anexo 3.10: Grafica de la frecuencia del modo de preparación de los estudiantes para sus clases.**



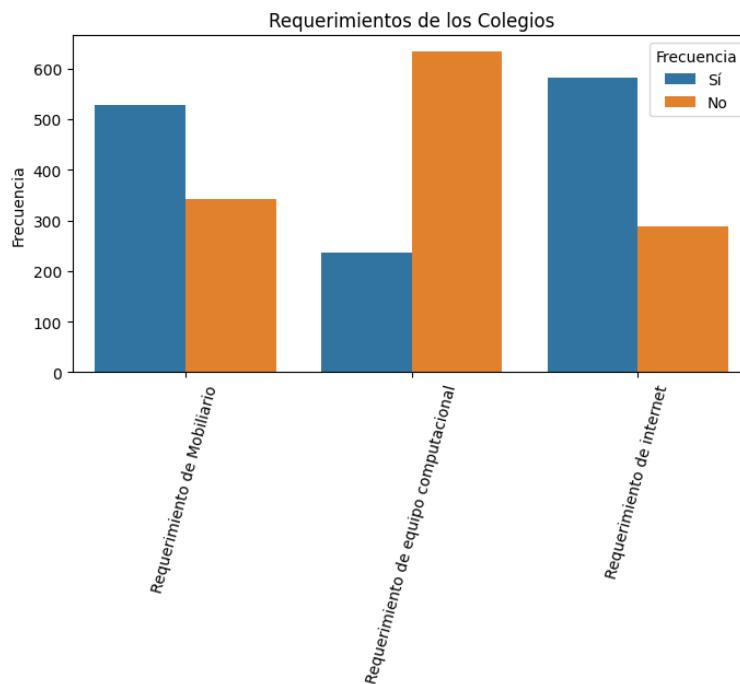
**Anexo 3.11: Grafica de frecuencia de elaboración del material de enseñanza de los profesores.**



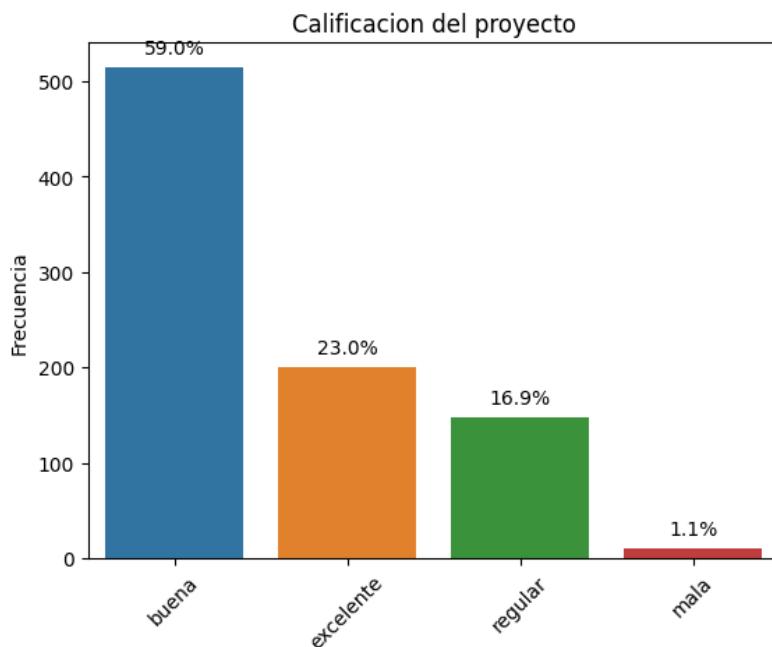
Anexo 3.12: Grafica de la frecuencia de la modalidad de Clases en las enseñanzas de los profesores.



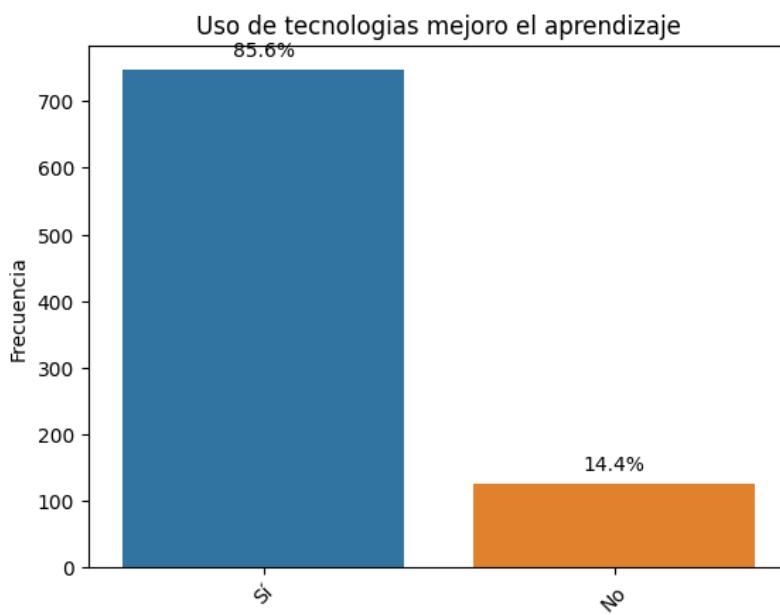
Anexo 3.13: Grafica de la frecuencia de los requerimientos de infraestructura de las escuelas públicas de Bolivia.



**Anexo 3.14: Grafica de la calificación que tuvo el proyecto de inclusión digital.**



**Anexo 3.15: Grafica de la frecuencia que determina si el uso de tecnología mejoró el aprendizaje en los estudiantes.**



## Anexo 4: Algoritmo Logistic Regression como Modelo de clasificación.

```

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt

# Variables predictoras
X = df_encoded[['Departamento','internet en colegio_Sí', "Aula en colegio para uso de computadoras_Sí",
                 "Modalidad de Clases_Clases de laboratorios",'Modalidad de Clases_Clases prácticas',
                 'Incorporacion de TIC en aula_Casi siempre','Incorporacion de TIC en aula_Ocasionalmente',
                 'Incorporacion de TIC en aula_Rara vez','Incorporacion de TIC en aula_Siempre',
                 'Uso de tecnologias mejoro el aprendizaje_Sí']]

# Variable objetivo
y = df_encoded["impacto de inclusion digital_Alto"]

# Dividision de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Creacion del modelo de regresión logística
model = LogisticRegression(max_iter=100)

#Entrenamiento del modelo
model.fit(X_train, y_train)

#Prueba del modelo
y_pred = model.predict(X_test)

# Calculo de las métricas de evaluación
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
y_prob = model.predict_proba(X_test)[:, 1]
roc_auc = roc_auc_score(y_test, y_prob)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred)

# Impresion de las métricas
print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1-score:", f1)
print("AUC-ROC:", roc_auc)

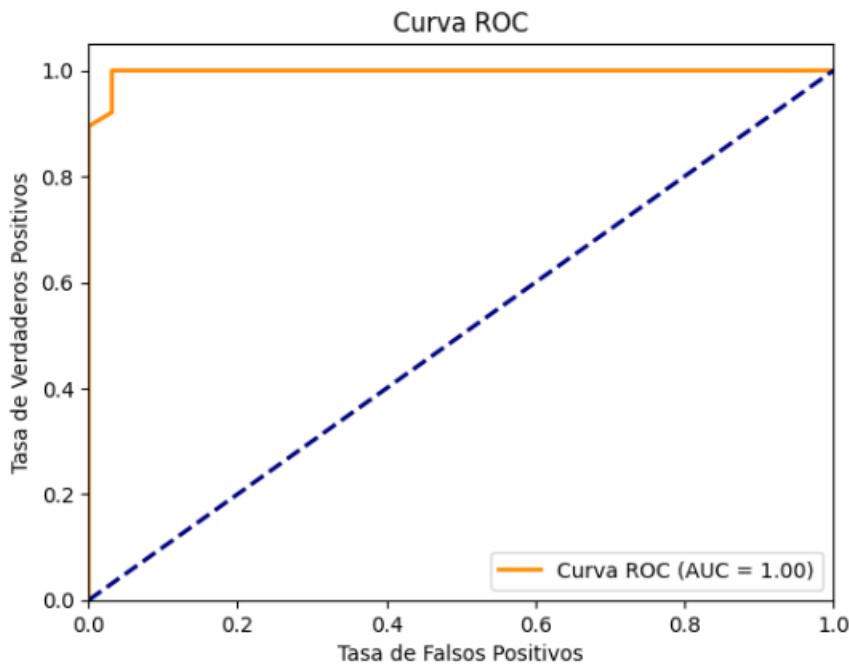
# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)

```

```
# Curva ROC
fpr, tpr, thresholds = roc_curve(y_test, y_prob)
plt.figure()
plt.plot(fpr, tpr, color='darkorange', lw=2, label='Curva ROC (AUC = %0.2f)' % roc_auc)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de Falsos Positivos')
plt.ylabel('Tasa de Verdaderos Positivos')
plt.title('Curva ROC')
plt.legend(loc="lower right")
plt.show()

Accuracy: 0.9714285714285714
Precision: 0.95
Recall: 1.0
F1-score: 0.9743589743589743
AUC-ROC: 0.9971217105263158

Matriz de Confusión:
[[30  2]
 [ 0 38]]
```



Link:

[https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive\\_link](https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive_link)

## Anexo 5: Algoritmo Gaussian como modelo de clasificación.

```

from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt

# Variables predictoras
X = df_encoded[['Departamento', 'internet en colegio_Sí', 'Aula en colegio para uso de computadoras_Sí',
                 'Modalidad de Clases_Clases de laboratorios', 'Modalidad de Clases_Clases prácticas',
                 'Incorporacion de TIC en aula_Casi siempre',
                 'Incorporacion de TIC en aula_Ocasionalmente', 'Incorporacion de TIC en aula_Rara vez',
                 'Incorporacion de TIC en aula_Siempre', 'Uso de tecnologias mejoró el aprendizaje_Sí', ]]

# Variable onjetivo
y = df_encoded["impacto de inclusion digital_Alto"]

# Dividision de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Crear del Modelo
nb_model = GaussianNB()

#Entrenamiento del modelo
nb_model.fit(X_train, y_train)

#Predicciones en el conjunto de prueba
y_pred_nb = nb_model.predict(X_test)

# Calculo de métricas de evaluación
accuracy_nb = accuracy_score(y_test, y_pred_nb)
precision_nb = precision_score(y_test, y_pred_nb)
recall_nb = recall_score(y_test, y_pred_nb)
f1_nb = f1_score(y_test, y_pred_nb)
y_prob_nb = nb_model.predict_proba(X_test)[:, 1]
roc_auc_nb = roc_auc_score(y_test, y_prob_nb)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred_nb)

# Impresion de métricas
print("Accuracy (Naive Bayes):", accuracy_nb)
print("Precision (Naive Bayes):", precision_nb)
print("Recall (Naive Bayes):", recall_nb)
print("F1-score (Naive Bayes):", f1_nb)
print("AUC-ROC (Naive Bayes):", roc_auc_nb)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)

```

```
# Curva ROC
fpr_nb, tpr_nb, thresholds_nb = roc_curve(y_test, y_prob_nb)
plt.figure()
plt.plot(fpr_nb, tpr_nb, color='darkorange', lw=2, label='Curva ROC (AUC = %0.2f)' % roc_auc_nb)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de Falsos Positivos')
plt.ylabel('Tasa de Verdaderos Positivos')
plt.title('Curva ROC (Naive Bayes)')
plt.legend(loc="lower right")
plt.show()
```

Accuracy (Naive Bayes): 0.9714285714285714

Precision (Naive Bayes): 0.95

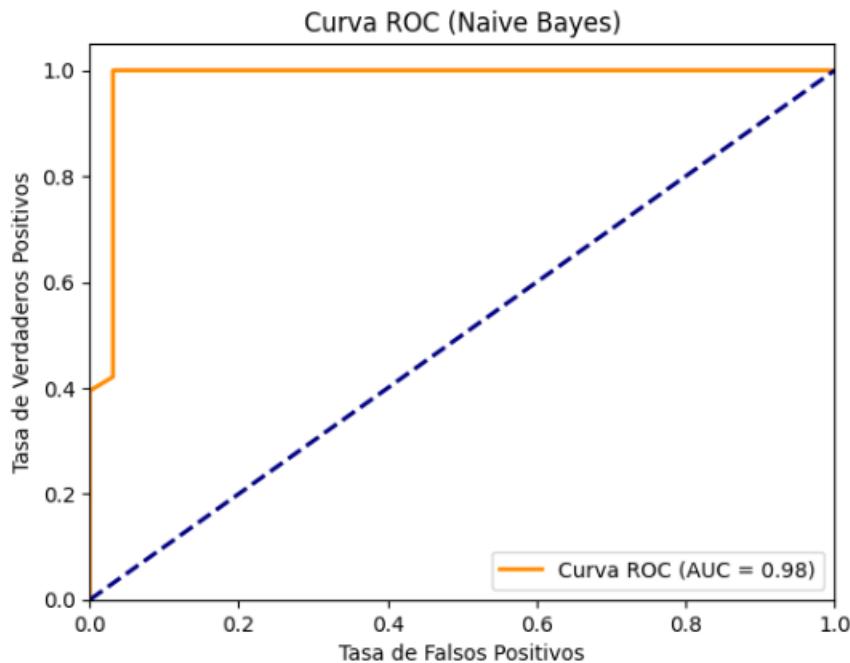
Recall (Naive Bayes): 1.0

F1-score (Naive Bayes): 0.9743589743589743

AUC-ROC (Naive Bayes): 0.9814967105263158

Matriz de Confusión:

```
[[30  2]
 [ 0 38]]
```



Link:

[https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive\\_link](https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive_link)

## Anexo 6: Algoritmo KNeighbors como modelo de clasificación.

```

from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt

#Variables predictoras
X = df_encoded[['Departamento','internet en colegio_Sí', 'Aula en colegio para uso de computadoras_Sí',
                 'Modalidad de Clases_Clases de laboratorios','Modalidad de Clases_Clases prácticas',
                 'Incorporacion de TIC en aula_Casi siempre','Incorporacion de TIC en aula_Ocasionalmente',
                 'Incorporacion de TIC en aula_Rara vez','Incorporacion de TIC en aula_Siempre',
                 'Uso de tecnologias mejoro el aprendizaje_Sí',]]

#Variable objetivo
y = df_encoded["impacto de inclusion digital_Alto"]

# Dividision de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Creacion del modelo de k-Nearest Neighbors (k-NN)
knn_model = KNeighborsClassifier(n_neighbors=3)

#Entrenamiento del modelo
knn_model.fit(X_train, y_train)

#Predicciones del conjunto de prueba
y_pred_knn = knn_model.predict(X_test)

# Calculo de métricas de evaluación
accuracy_knn = accuracy_score(y_test, y_pred_knn)
precision_knn = precision_score(y_test, y_pred_knn)
recall_knn = recall_score(y_test, y_pred_knn)
f1_knn = f1_score(y_test, y_pred_knn)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred_knn)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)

# No se calcula AUC-ROC para k-NN ya que no genera probabilidades directas

# Curva ROC (No aplicable para k-NN)
# k-NN no genera probabilidades directas, por lo que no se puede calcular una curva ROC en este caso.

```

Accuracy (k-NN): 0.8857142857142857  
Precision (k-NN): 0.8571428571428571  
Recall (k-NN): 0.9473684210526315  
F1-score (k-NN): 0.9

Matriz de Confusión:  
[[26 6]  
 [ 2 36]]

Link:

[https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive\\_link](https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive_link)

## Anexo 7: Modelo de clasificación del algoritmo SVM o SVC.

```
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt

#Variables predictoras
X = df_encoded[['Departamento','internet en colegio_Sí', "Aula en colegio para uso de computadoras_Sí",
                 "Modalidad de Clases_Clases de laboratorios",'Modalidad de Clases_Clases prácticas',
                 'Incorporacion de TIC en aula_Casi siempre','Incorporacion de TIC en aula_Ocasionalmente',
                 'Incorporacion de TIC en aula_Rara vez','Incorporacion de TIC en aula_Siempre',
                 'Uso de tecnologias mejoro el aprendizaje_Sí',]]]

#Variable objetivo
y = df_encoded["impacto de inclusion digital_Alto"]

# Dividision de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Creacion del modelo SVM con kernel radial (RBF)
svm_model = SVC(kernel='rbf', random_state=42, probability=True)

#Entrenamiento del modelo
svm_model.fit(X_train, y_train)

#Predicciones del conjunto de prueba
y_pred_svm = svm_model.predict(X_test)

# Calculo de métricas de evaluación
accuracy_svm = accuracy_score(y_test, y_pred_svm)
precision_svm = precision_score(y_test, y_pred_svm)
recall_svm = recall_score(y_test, y_pred_svm)
f1_svm = f1_score(y_test, y_pred_svm)
y_prob_svm = svm_model.predict_proba(X_test)[:, 1]
roc_auc_svm = roc_auc_score(y_test, y_prob_svm)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred_svm)

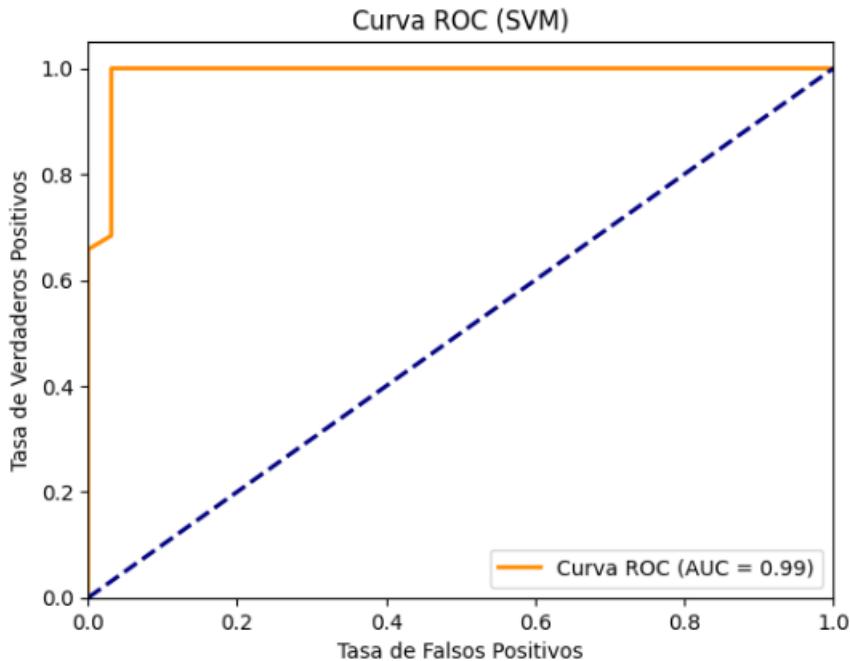
# Impresion de las métricas
print("Accuracy (SVM):", accuracy_svm)
print("Precision (SVM):", precision_svm)
print("Recall (SVM):", recall_svm)
print("F1-score (SVM):", f1_svm)
print("AUC-ROC (SVM):", roc_auc_svm)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)
```

```
# Curva ROC
fpr_svm, tpr_svm, thresholds_svm = roc_curve(y_test, y_prob_svm)
plt.figure()
plt.plot(fpr_svm, tpr_svm, color='darkorange', lw=2, label='Curva ROC (AUC = %0.2f)' % roc_auc_svm)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de Falsos Positivos')
plt.ylabel('Tasa de Verdaderos Positivos')
plt.title('Curva ROC (SVM)')
plt.legend(loc="lower right")
plt.show()
```

Accuracy (SVM): 0.9428571428571428  
Precision (SVM): 0.9047619047619048  
Recall (SVM): 1.0  
F1-score (SVM): 0.9500000000000001  
AUC-ROC (SVM): 0.989720394736842

Matriz de Confusión:  
[[28 4]  
 [ 0 38]]



Link:

[https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive\\_link](https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive_link)

## Anexo 8: Modelo de clasificación del algoritmo Redes Neuronales ANN.

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from tensorflow import keras
from tensorflow.keras import layers
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt

# Variables predictoras
X = df_encoded[['Departamento','internet en colegio_Sí', "Aula en colegio para uso de computadoras_Sí",
                 "Modalidad de Clases_Clases de laboratorios",'Modalidad de Clases_Clases prácticas',
                 'Incorporacion de TIC en aula_Casi siempre','Incorporacion de TIC en aula_Ocasionalmente',
                 'Incorporacion de TIC en aula_Rara vez','Incorporacion de TIC en aula_Siempre',
                 'Uso de tecnologias mejoro el aprendizaje_Sí']]

# Variables objetivo
y = df_encoded["impacto de inclusion digital_Alto"]

# División de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Escalamiento de las características para mejorar el rendimiento de la red neuronal
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Creación de un modelo de red neuronal artificial (ANN)
model = keras.Sequential([
    layers.Input(shape=(X_train_scaled.shape[1],)), # Capa de entrada con el número de características
    layers.Dense(64, activation='relu'), # Capa oculta con 64 neuronas y función de activación ReLU
    layers.Dense(1, activation='sigmoid') # Capa de salida con 1 neurona y función de
                                         #activación sigmoide (para clasificación binaria)
])

# Compilación del modelo
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Entrenamiento el modelo
model.fit(X_train_scaled, y_train, epochs=100, batch_size=64, validation_split=0.2)

# Predicciones en el conjunto de prueba
y_prob_ann = model.predict(X_test_scaled)
y_pred_ann = (y_prob_ann > 0.5).astype(int)

# Calculo de métricas de evaluación
accuracy_ann = accuracy_score(y_test, y_pred_ann)
precision_ann = precision_score(y_test, y_pred_ann)
recall_ann = recall_score(y_test, y_pred_ann)
f1_ann = f1_score(y_test, y_pred_ann)
roc_auc_ann = roc_auc_score(y_test, y_prob_ann)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred_ann)

# Impresión de las métricas
print("Accuracy (ANN):", accuracy_ann)
print("Precision (ANN):", precision_ann)
print("Recall (ANN):", recall_ann)
print("F1-score (ANN):", f1_ann)
print("AUC-ROC (ANN):", roc_auc_ann)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)
```

```

# Curva ROC
fpr_ann, tpr_ann, thresholds_ann = roc_curve(y_test, y_prob_ann)
plt.figure()
plt.plot(fpr_ann, tpr_ann, color='darkorange', lw=2, label='Curva ROC (AUC = %.2f)' % roc_auc_ann)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de Falsos Positivos')
plt.ylabel('Tasa de Verdaderos Positivos')
plt.title('Curva ROC (ANN)')
plt.legend(loc="lower right")
plt.show()

4/4 [=====] - 0s 18ms/step - loss: 0.0679 - accuracy: 0.9818 - val_loss: 0.1134 - val_accuracy: 0.9643
Epoch 91/100
4/4 [=====] - 0s 12ms/step - loss: 0.0671 - accuracy: 0.9818 - val_loss: 0.1130 - val_accuracy: 0.9643
Epoch 92/100
4/4 [=====] - 0s 12ms/step - loss: 0.0663 - accuracy: 0.9818 - val_loss: 0.1123 - val_accuracy: 0.9643
Epoch 93/100
4/4 [=====] - 0s 12ms/step - loss: 0.0657 - accuracy: 0.9818 - val_loss: 0.1118 - val_accuracy: 0.9643
Epoch 94/100
4/4 [=====] - 0s 19ms/step - loss: 0.0649 - accuracy: 0.9818 - val_loss: 0.1112 - val_accuracy: 0.9643
Epoch 95/100
4/4 [=====] - 0s 13ms/step - loss: 0.0643 - accuracy: 0.9818 - val_loss: 0.1105 - val_accuracy: 0.9643
Epoch 96/100
4/4 [=====] - 0s 15ms/step - loss: 0.0636 - accuracy: 0.9818 - val_loss: 0.1099 - val_accuracy: 0.9643
Epoch 97/100
4/4 [=====] - 0s 12ms/step - loss: 0.0630 - accuracy: 0.9818 - val_loss: 0.1093 - val_accuracy: 0.9643
Epoch 98/100
4/4 [=====] - 0s 14ms/step - loss: 0.0624 - accuracy: 0.9818 - val_loss: 0.1090 - val_accuracy: 0.9643
Epoch 99/100
4/4 [=====] - 0s 13ms/step - loss: 0.0617 - accuracy: 0.9818 - val_loss: 0.1085 - val_accuracy: 0.9643
Epoch 100/100
4/4 [=====] - 0s 11ms/step - loss: 0.0612 - accuracy: 0.9818 - val_loss: 0.1081 - val_accuracy: 0.9643
3/3 [=====] - 0s 3ms/step

Accuracy (ANN): 0.9857142857142858
Precision (ANN): 0.9743589743589743
Recall (ANN): 1.0
F1-score (ANN): 0.9870129870129869
AUC-ROC (ANN): 0.9979440789473684

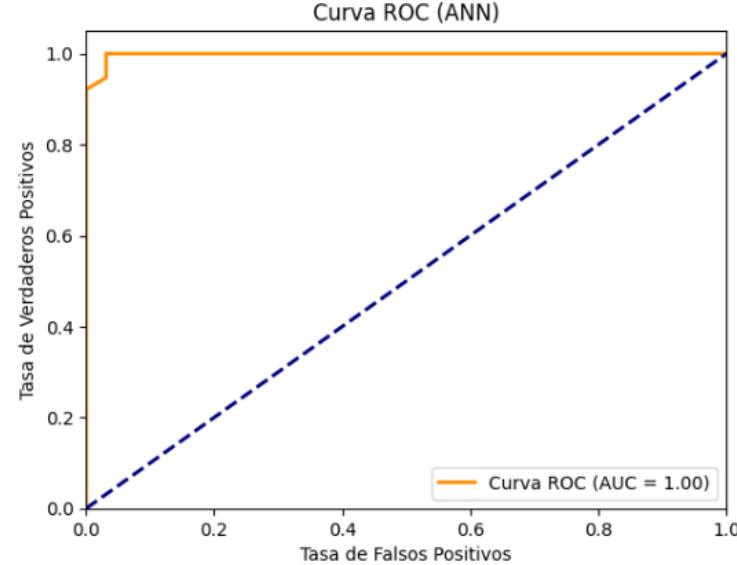
```

Matriz de Confusión:

```

[[31  1]
 [ 0 38]]

```



Link:

[https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive\\_link](https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive_link)

## Anexo 9: Algoritmo Decision Tree como modelo de clasificación.

```
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt

#Variable predictora
X = df_encoded[['Departamento','internet en colegio_Sí', "Aula en colegio para uso de computadoras_Sí",
                 "Modalidad de Clases_Clases de laboratorios",'Modalidad de Clases_Clases prácticas',
                 'Incorporacion de TIC en aula_Casi siempre','Incorporacion de TIC en aula_Ocasionalmente',
                 'Incorporacion de TIC en aula_Rara vez','Incorporacion de TIC en aula_Siempre',
                 'Uso de tecnologias mejoro el aprendizaje_Sí']]]

#Variable objetivo
y = df_encoded["impacto de inclusion digital_Alto"]

# Division de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Crear un modelo de Árboles de Decisión
tree_model = DecisionTreeClassifier(random_state=42)

#Entrenamiento del modelo
tree_model.fit(X_train, y_train)

#Predicciones en el conjunto de prueba
y_prob_tree = tree_model.predict_proba(X_test)[:, 1]
y_pred_tree = (y_prob_tree > 0.5).astype(int)

# Calculo de métricas de evaluación
accuracy_tree = accuracy_score(y_test, y_pred_tree)
precision_tree = precision_score(y_test, y_pred_tree)
recall_tree = recall_score(y_test, y_pred_tree)
f1_tree = f1_score(y_test, y_pred_tree)
roc_auc_tree = roc_auc_score(y_test, y_prob_tree)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred_tree)

# Impresion de las métricas
print("Accuracy (Decision Tree):", accuracy_tree)
print("Precision (Decision Tree):", precision_tree)
print("Recall (Decision Tree):", recall_tree)
print("F1-score (Decision Tree):", f1_tree)
print("AUC-ROC (Decision Tree):", roc_auc_tree)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)
```

```
# Curva ROC
fpr_tree, tpr_tree, thresholds_tree = roc_curve(y_test, y_prob_tree)
plt.figure()
plt.plot(fpr_tree, tpr_tree, color='darkorange', lw=2, label='Curva ROC (AUC = %.2f)' % roc_auc_tree)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de Falsos Positivos')
plt.ylabel('Tasa de Verdaderos Positivos')
plt.title('Curva ROC (Decision Tree)')
plt.legend(loc="lower right")
plt.show()
```

Accuracy (Decision Tree): 0.9571428571428572

Precision (Decision Tree): 0.972972972972973

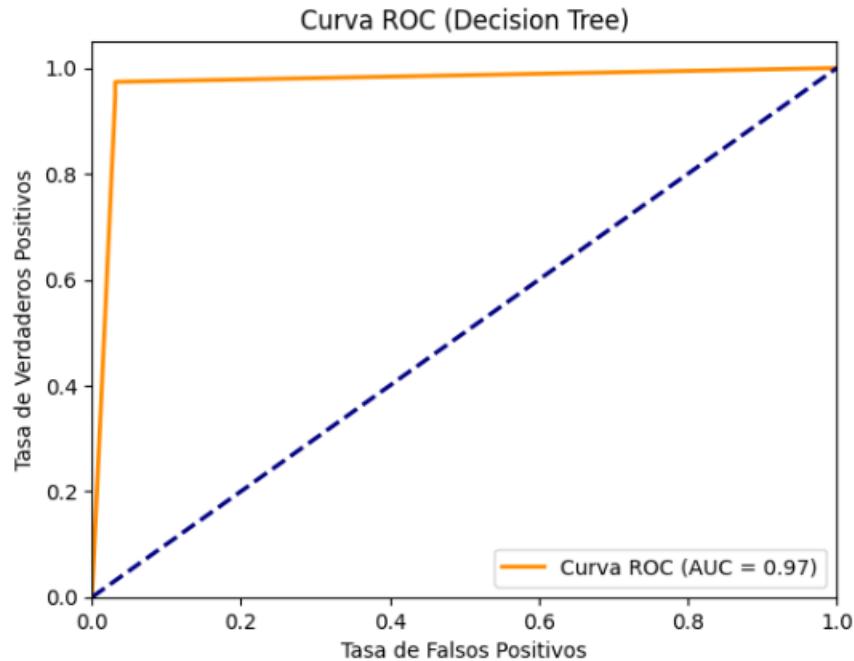
Recall (Decision Tree): 0.9473684210526315

F1-score (Decision Tree): 0.9599999999999999

AUC-ROC (Decision Tree): 0.9708059210526316

Matriz de Confusión:

31	1
2	36



Link:

[https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive\\_link](https://drive.google.com/file/d/1b7cUcxEmpD114gxy3ZOhsqA2lDK0eito/view?usp=drive_link)

## Anexo 10: Algoritmo Redes Neuronales DNN como modelo de clasificación.

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from tensorflow import keras
from tensorflow.keras import layers
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score, roc_curve
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt

#Variable predictora
X = df_encoded[['Departamento','internet en colegio_Sí', "Aula en colegio para uso de computadoras_Sí",
                 'Modalidad de Clases_Clases de laboratorios','Modalidad de Clases_Clases prácticas',
                 'Incorporacion de TIC en aula_Casi siempre','Incorporacion de TIC en aula_Ocasionalmente',
                 'Incorporacion de TIC en aula_Rara vez','Incorporacion de TIC en aula_Siempre',
                 'Uso de tecnologias mejoro el aprendizaje_Sí',]]]

#Variable objetivo
y = df_encoded["impacto de inclusion digital_Alto"]

# Division de los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Escalamiento de las características para normalizarlas
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Creacion de un modelo de red neuronal profunda (DNN)
model = keras.Sequential([
    layers.Input(shape=(X_train_scaled.shape[1],)),
    layers.Dense(128, activation='tanh'),
    layers.Dropout(0.2),
    layers.Dense(64, activation='tanh'),
    layers.Dropout(0.2),
    layers.Dense(1, activation='tanh')
])

# Compilacion del modelo
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Entrenamiento del modelo
history = model.fit(X_train_scaled, y_train, epochs=100, batch_size=32, validation_split=0.2, verbose=1)

#Predicciones en el conjunto de prueba
y_prob_dnn = model.predict(X_test_scaled)
y_pred_dnn = (y_prob_dnn > 0.5).astype(int)

# Calculo de métricas de evaluación
accuracy_dnn = accuracy_score(y_test, y_pred_dnn)
precision_dnn = precision_score(y_test, y_pred_dnn)
recall_dnn = recall_score(y_test, y_pred_dnn)
f1_dnn = f1_score(y_test, y_pred_dnn)
roc_auc_dnn = roc_auc_score(y_test, y_prob_dnn)

# Matriz de confusión
conf_matrix = confusion_matrix(y_test, y_pred_dnn)
```

```

# Impresión de las métricas
print("Accuracy (Deep Neural Network):", accuracy_dnn)
print("Precision (Deep Neural Network):", precision_dnn)
print("Recall (Deep Neural Network):", recall_dnn)
print("F1-score (Deep Neural Network):", f1_dnn)
print("AUC-ROC (Deep Neural Network):", roc_auc_dnn)

# Matriz de confusión
print("\nMatriz de Confusión:")
print(conf_matrix)

# Curva ROC
fpr_dnn, tpr_dnn, thresholds_dnn = roc_curve(y_test, y_prob_dnn)
plt.figure()
plt.plot(fpr_dnn, tpr_dnn, color='darkorange', lw=2, label='Curva ROC (AUC = %0.2f)' % roc_auc_dnn)
plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('Tasa de Falsos Positivos')
plt.ylabel('Tasa de Verdaderos Positivos')
plt.title('Curva ROC (Deep Neural Network)')
plt.legend(loc="lower right")
plt.show()

Epoch 1/100
7/7 [=====] - 1s 38ms/step - loss: 3.1426 - accuracy: 0.5773 - val_loss: 1.3955 - val_accuracy: 0.8571
Epoch 2/100
7/7 [=====] - 0s 11ms/step - loss: 1.5654 - accuracy: 0.7909 - val_loss: 0.6289 - val_accuracy: 0.8393
Epoch 3/100
7/7 [=====] - 0s 10ms/step - loss: 0.7410 - accuracy: 0.8318 - val_loss: 0.5500 - val_accuracy: 0.8929
Epoch 4/100
7/7 [=====] - 0s 8ms/step - loss: 0.4836 - accuracy: 0.8727 - val_loss: 0.2405 - val_accuracy: 0.8929
Epoch 5/100
7/7 [=====] - 0s 10ms/step - loss: 0.3229 - accuracy: 0.9318 - val_loss: 0.1926 - val_accuracy: 0.9464
Epoch 6/100
    Accuracy (Deep Neural Network): 0.9857142857142858
    Precision (Deep Neural Network): 0.9743589743589743
    Recall (Deep Neural Network): 1.0
    F1-score (Deep Neural Network): 0.9870129870129869
    AUC-ROC (Deep Neural Network): 0.9995888157894737

```

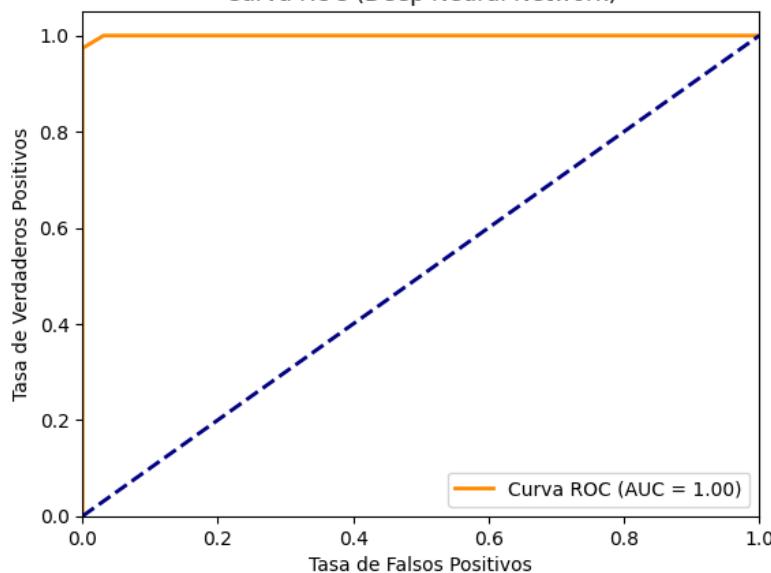
Matriz de Confusión:

```

[[31  1]
 [ 0 38]]

```

Curva ROC (Deep Neural Network)



**Anexo 11: Encuesta a los profesores, para ayuda a definir variables de clasificación.**

¿Cuál es el nombre de la unidad educativa en la que enseña?	¿Cuántos años de experiencia tiene?	¿Qué materias imparte en su colegio?	¿Su colegio cuenta con servicio a internet?	¿Su colegio cuenta con un aula de computación?	Las clases que imparte en aula son:	Permite el uso de celular en clase con:	Mencione que red social utiliza con frecuencia para fines académicos [Ejemplo: Facebook, WhatsApp, Telegram, Twitter]
Gerardo Vargas Vidal	32	5 años	Biology	Sí	herramientas como pizarrones inteligentes, Práctico	Sí	WhatsApp
Americano Vecinal C	36	15 años	Primaria	No	herramientas como pizarrones inteligentes, Teórico;Práctico No	No	WhatsApp
SUECIA	38	8	BIOLOGÍA	Sí	Bibliotecas digital	Si	WhatsApp
Suecia	59	29	Valores, Espiritualidad y Religiones	Sí	herramientas como pizarrones inteligentes, Teórico;Práctico Si	Si	WhatsApp, Google
Prof .Jaime Alfonso Escalante G	36	10	Comunicación y Lenguajes	No	Computadoras para los estudiantes,Acceso Teórico	No	Facebook, tik tok
Suecia	37	11	Matemática, biología	Sí	Acceso a internet en todas las aulas,"herram"	Si	WhatsApp
Americano Vecinal C	58	35	Todas	No	Computadoras para los estudiantes,Aula cc Teórico;Práctico No	No	WhatsApp
Suecia	40	18	Educación Musical	Sí	herramientas como pizarrones inteligentes, Teórico;Práctico Si	Si	Facebook, WhatsApp, telegram, etc.
Suecia	40	23	Comunicación y Lenguajes	Sí	Computadoras para los estudiantes,Acceso Teórico;Práctico Si	Si	Facebook, WhatsApp, Tik Tok
Suecia	53	30	Enseña Tecnologica General	Sí	herramientas como pizarrones inteligentes, Teórico;Práctico Si	Si	Facebook, WhatsApp
SUECIA	36	13	CIENCIAS SOCIALES	Sí	Acceso a internet en todas las aulas,"herram"	Si	FACEBOOK, WHATSAPP
U. E. Suecia	33	10	Artes plásticas y visuales	No	Computadoras para los estudiantes,Aula cc Teórico;Práctico Si	Si	Facebook
Bolivia A	58	37	Literatura y lenguajes	Sí	Acceso a internet en todas las aulas,"herram" Teórico;Práctico	Si	Facebook
Unidad Educativa Americano	40	8 años	Computación	Sí	herramientas como pizarrones inteligentes, Teórico;Práctico No	No	WhatsApp
San Francisco II	65	30	Matemática	No	Acceso a internet en todas las aulas.	Práctico	No
Nuestra Señor del Pilar	40	10	Filosofía	Sí	Acceso a internet en todas las aulas,"bibliot	Si	WhatsApp
Nuestra Señora del Pilar	47	23	Sociales	Sí	Acceso a internet en todas las aulas,"bibliot"	Si	Telegram
Nuestra Señora Del Pilar	50	26 años	Química y Biología	Sí	Bibliotecas digital	Si	WhatsApp
Vecinal C	70	37	Primaria EFI	No	"Aula computacional."	Práctico	No
Americano vecinal C	55	30	Primeros	Sí	"Aula computacional."	Teórico;Práctico No	Facebook
Americano Vecinal C	43	21	Lengua y Ciencias Sociales	No	Computadoras para los estudiantes,Acceso Teórico;Práctico No	Si	WhatsApp
Americano Vecinal C	50	15	Primaria	No	Computadoras para los estudiantes,Aula cc Teórico;Práctico No	No	You tube
Americano Vecinal C	51	23	Primaria	No	Computadoras para los estudiantes,Aula cc Teórico	No	WhatsApp
Nuestra Señora del Pilar	26	18	Biología y Química	Sí	herramientas como pizarrones inteligentes, Teórico;Práctico Si	Si	WhatsApp, Telegram

## Anexo Principal CD.

Toda la información del presente proyecto se encuentra en el cd que se muestra continuación, con la estructura de carpetas siguientes:



En esta carpeta se encuentra 5 subcarpetas con toda la información respecto al presente proyecto.



Código de Modelo



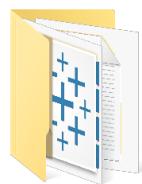
Dataset



Encuesta para ayuda de clasificación



Monografía



Visualización Tableau

En la carpeta de nombre “Código de Modelo”, se encuentra toda la parte de programación realizada en Python en Google Colab.

En la carpeta de nombre “Dataset”, se encuentra el conjunto de datos usados para la creación del modelo de Clasificación con aprendizaje supervisado. Se encuentra en un archivo de tipo Excel con nombre “profesores\_final”

En la carpeta de nombre “Encuesta para ayuda de clasificación”, se encuentra la información sobre la encuesta realizada a los 24 profesores en el departamento de Cochabamba para poder definir el método de clasificación del impacto del programa inclusión digital.

En la carpeta de nombre “monografía” se encuentra el presente documento en formato pdf.

En la carpeta de nombre “Visualización Tableau”, se encuentra el archivo de visualizaciones del conjunto de datos estudiados.

Link:

[https://drive.google.com/drive/folders/1MG5GzKTrh8NeaobP2EpwqSaRGXzEqCpE?usp=drive\\_link](https://drive.google.com/drive/folders/1MG5GzKTrh8NeaobP2EpwqSaRGXzEqCpE?usp=drive_link)

Para poder hacer correr el código en Colab:

- 1: Vincular tu Drive con google Colab
- 2: Subir el archivo "profesores\_final.xlsx" en la subcarpeta MyDrive de la carpeta drive "drive/MyDrive/profesores\_final.xlsx"
- 3: Subir el archivo "dataset4.xlsx" en la subcarpeta MyDrive de la carpeta drive "drive/MyDrive/dataset4.xlsx"

Para ver las visualizaciones debes:

- 1: instalar Power Bi en tu computadora.
- 2: abrir el archivo "INCLUSION DIGITAL.pbix"
- 3: es todo.

Para poder ver las visualizaciones de Tableau:

- 1: Crear en disco local D una carpeta llamada "DIPLOMADO ESTADISTICA", y dentro de esta carpeta crear una carpeta llamada "proyecto", y pegar ahí el archivo excel llamado "profesores\_fin - copia - copia.xlsx"
- 2: instalar Tableau.
- 3: Abrir el programa "PROYECTO\_DIPLOMADO".
- 4: es todo.