**TALENTO TECH**

**INTELIGENCIA ARTIFICIAL (IA)**

**NIVEL EXPLIRADOR**

CATE-IA: PLATAFORMA INTERACTIVA PARA EL APRENDIZAJE CATEQUÉTICO DE NIÑOS Y

JÓVENES

**Proyecto de Machine Learning**

Diócesis de Caldas, Antioquia

**INTEGRANTES:**

Doris María Vásquez González José Fernando Gómez Valderrama Juan Camilo Cano Velásquez

Unilasallista

Medellín, Colombia

2025

**TABLA DE CONTENIDOS**

[1. RESUMEN EJECUTIVO 4](#_Toc208946175)

[2. DEFINICIÓN DEL PROBLEMA 5](#_Toc208946176)

[2.1 Descripción del Problema 5](#_Toc208946177)

[2.2 Objetivos del Proyecto 5](#_Toc208946178)

[2.3 Pregunta de Investigación 6](#_Toc208946179)

[3. MARCO TEÓRICO Y JUSTIFICACIÓN 7](#_Toc208946180)

[3.1 Contexto Religioso 8](#_Toc208946181)

[3.2 Contexto Tecnológico y Social 8](#_Toc208946182)

[3.3 Fundamentos de Machine Learning en Educación 8](#_Toc208946183)

[3.4 Oportunidad Pedagógica 9](#_Toc208946184)

[3.5 Justificación del Proyecto 10](#_Toc208946185)

[4. METODOLOGÍA DE MACHINE LEARNING 11](#_Toc208946186)

[4.1 Enfoque Metodológico 11](#_Toc208946187)

[4.2 Estrategia de Modelado 11](#_Toc208946188)

[4.3 Métricas de Evaluación 12](#_Toc208946189)

[4.4 Validación y Optimización 12](#_Toc208946190)

[5. RECOLECCIÓN Y PREPARACIÓN DE DATOS 13](#_Toc208946191)

[5.1 Fuentes de Datos 13](#_Toc208946192)

[5.2 Diseño del Instrumento de Recolección 13](#_Toc208946193)

[5.3 Descripción del Dataset 13](#_Toc208946194)

[5.4 Preprocesamiento de Datos 14](#_Toc208946195)

[5.5 Calidad de los Datos 15](#_Toc208946196)

[6. ANÁLISIS EXPLORATORIO DE DATOS 16](#_Toc208946197)

[6.1 Análisis Descriptivo 16](#_Toc208946198)

[6.2 Análisis de Correlaciones 17](#_Toc208946199)

[6.3 Análisis de Distribuciones 17](#_Toc208946200)

[6.4 Análisis de Segmentación Inicial 18](#_Toc208946201)

[6.5 Insights Clave del EDA 18](#_Toc208946202)

[7. IMPLEMENTACIÓN DE MODELOS 19](#_Toc208946203)

[7.1 Regresión Lineal 19](#_Toc208946204)

[7.2 Regresión Logística 20](#_Toc208946205)

[7.3 K-Nearest Neighbors (KNN) 20](#_Toc208946206)

[7.4 K-Means Clustering 21](#_Toc208946207)

[7.5 Árboles de Decisión 22](#_Toc208946208)

[8. EVALUACIÓN Y COMPARACIÓN DE MODELOS 24](#_Toc208946209)

[8.1 Resumen Comparativo de Rendimiento 24](#_Toc208946210)

[8.2 Análisis Detallado por Modelo 25](#_Toc208946211)

[8.3 Validación Cruzada y Estabilidad 27](#_Toc208946212)

[8.4 Análisis de Overfitting/Underfitting 27](#_Toc208946213)

[8.5 Selección de Modelos por Caso de Uso 27](#_Toc208946214)

[8.6 Análisis de Complejidad Computacional 28](#_Toc208946215)

[9. INTERPRETACIÓN GRÁFICA DEL "CATE-IA" 29](#_Toc208946216)

[10. SEGMENTACIÓN DE USUARIOS 36](#_Toc208946217)

[10.1 Perfil de Segmentos Identificados 36](#_Toc208946218)

[11. INTERPRETACIÓN DE RESULTADOS 39](#_Toc208946219)

[10.1 Conexión entre Modelos y Hallazgos 39](#_Toc208946220)

[11. RECOMENDACIONES ESTRATÉGICAS 40](#_Toc208946221)

[11.2 Diseño Centrado en el Usuario y Personalización 40](#_Toc208946222)

[12. IMPLEMENTACIÓN Y DESPLIEGUE 41](#_Toc208946223)

[12.1 Fase 1: Desarrollo del Prototipo (2025) 41](#_Toc208946224)

[12.2 Fase 2: Pruebas Piloto y Evaluación (2025-2026) 41](#_Toc208946225)

[12.3 Fase 3: Despliegue y Escalabilidad (2026 en adelante) 41](#_Toc208946226)

[13. CONCLUSIONES 42](#_Toc208946227)

[14. TRABAJO FUTURO 43](#_Toc208946228)

[15. REFERENCIAS BIBLIOGRÁFICAS 44](#_Toc208946229)

[16. ANEXOS 46](#_Toc208946230)

# RESUMEN EJECUTIVO

El presente proyecto desarrolla "Cate-IA", una plataforma interactiva basada en Machine Learning para optimizar el aprendizaje catequético de niños y jóvenes de 7 a 14 años en la Diócesis de Caldas, Antioquia. La investigación analiza las preferencias digitales y hábitos de aprendizaje de 500 estudiantes para diseñar contenido educativo religioso personalizado y efectivo.

A través de la implementación de cinco modelos de Machine Learning (Regresión Lineal, Regresión Logística, K-Nearest Neighbors, K-Means Clustering y Árboles de Decisión), se identificaron patrones significativos en las preferencias de aprendizaje y se desarrolló una segmentación de usuarios que permite personalizar la experiencia catequética.

**Resultados Principales**

**Regresión Lineal:** R² = 0.85, RMSE = 0.52 en predicción de uso de internet para fe

**Regresión Logística:** 94% accuracy, ROC AUC = 0.99 en autorización de datos

**K-Nearest Neighbors:** 91% accuracy, F1 = 0.90 en clasificación de preferencias

**K-Means Clustering:** 4 clusters identificados, Silhouette = 0.65

**Árboles de Decisión:** 89% accuracy, Precisión = 0.88

La segmentación reveló cuatro perfiles distintos de usuarios: Digitales Avanzados (25%), Creativos Tradicionales (30%), Gamers Interactivos (20%) y Conservadores (25%), cada uno con necesidades y preferencias específicas que requieren estrategias diferenciadas de contenido.

Las recomendaciones estratégicas incluyen la priorización de contenido audiovisual, implementación de gamificación, desarrollo de avatares amigables y optimización para dispositivos móviles, con un enfoque de personalización basado en los segmentos identificados.

# DEFINICIÓN DEL PROBLEMA

## Descripción del Problema

¿Cómo la implementación de una plataforma digital (Cate-IA), que integra un chatbot y un avatar, impacta positivamente el nivel de comprensión y la participación de niños y jóvenes de 7 a 14 años en su preparación para la Primera Comunión y la Confirmación dentro de la Diócesis de Caldas, Antioquia, durante el período 2025- 2026?

La Iglesia Católica enfrenta el desafío de conectar con las nuevas generaciones, que se desarrollan en entornos altamente digitales, visuales y participativos. Los métodos tradicionales de catequesis muestran una desconexión creciente con el lenguaje, intereses y formas de aprender de los niños y jóvenes actuales.

## Objetivos del Proyecto

* + 1. **Objetivo General**

Crear una plataforma interactiva para el aprendizaje catequético de niños y jóvenes en su preparación a los sacramentos de Primera Comunión y Confirmación, en la Diócesis del municipio de Caldas, Antioquia.

* + 1. **Objetivos Específicos**
       1. Analizar el currículo de catequesis de la Diócesis de Caldas, Antioquia, para identificar los temas clave y las necesidades de los usuarios.
       2. Diseñar un guión interactivo y un avatar amigable que se integren en un chatbot, asegurando que el lenguaje sea accesible y atractivo para el público objetivo.
       3. Desarrollar un prototipo funcional del "Cate-IA" con una arquitectura full stack que permite la interacción con el usuario y la gestión del contenido.
       4. Evaluar la efectividad del "Cate-IA" en la retención de conocimiento y la motivación de los participantes mediante la aplicación de pruebas piloto con grupos focales de catequistas y niños.
       5. Proponer un plan de implementación y escalabilidad para la adopción de la herramienta en toda la Diócesis de Caldas.

## Pregunta de Investigación

La investigación busca responder cómo los algoritmos de Machine Learning pueden identificar patrones en las preferencias de aprendizaje de los niños y jóvenes para crear experiencias catequéticas personalizadas y más efectivas, manteniendo la profundidad espiritual y doctrinal de la formación sacramental.

# MARCO TEÓRICO Y JUSTIFICACIÓN

La integración de la inteligencia artificial (IA) y, específicamente, de técnicas de machine learning en la educación religiosa y la catequesis digital representa un campo emergente que busca responder a los desafíos pedagógicos de las nuevas generaciones inmersas en entornos digitales. Este abordaje tecnológico adquiere relevancia pues los métodos tradicionales de catequesis muestran una desconexión creciente con las formas actuales de aprendizaje basadas en dispositivos móviles, videojuegos y plataformas interactivas.

Diversos estudios muestran cómo la IA puede personalizar el proceso educativo adaptando contenidos y estrategias pedagógicas según las necesidades y preferencias individuales de los estudiantes, incrementando así la motivación, la participación y la retención del aprendizaje (Estrada, 2024; Martínez et al., 2023). La aplicación de machine learning en contextos educativos permite identificar patrones de comportamiento y segmentar a los usuarios según sus estilos de aprendizaje, facilitando la creación de experiencias catequéticas personalizadas (Papakostas, 2025; Rahman, 2024).

En el ámbito religioso, la IA se ha explorado como herramienta para la evangelización y el acompañamiento espiritual, ofreciendo nuevas vías de comunicación más alineadas con la cultura digital. En este sentido, la gamificación se destaca como una estrategia clave para transformar la catequesis en una experiencia atractiva y significativa, aumentando el compromiso de niños y jóvenes a través de dinámicas de juego y sistemas de recompensas (Rivera Rojas, 2024; Velasco, 2021).

Asimismo, investigaciones recientes subrayan la importancia de que la adopción de estas tecnologías se realice bajo principios éticos sólidos que respeten la dignidad humana y la privacidad, evitando sesgos algorítmicos y promoviendo la equidad educativa (Vatican, 2025; Giraldo, 2011). La implementación cuidadosa y estratégica de plataformas digitales gamificadas en la catequesis permite mantener la profundidad espiritual y doctrinal mientras se adapta el formato al lenguaje y preferencias de los usuarios contemporáneos (Gutiérrez et al., 2019).

Finalmente, el despliegue de estas tecnologías en ambientes pastorales y educativos se considera una oportunidad para revalorizar la experiencia catequética, superando barreras tradicionales y contribuyendo a una evangelización contextualizada y efectiva en la era digital (Católicos en Red, 2023; IGLESIA de Concepción, 2025).

## Contexto Religioso

La Iglesia Católica tiene la misión de formar en la fe a los niños y jóvenes para recibir dignamente los sacramentos, especialmente la Primera Comunión y la Confirmación. Sin embargo, en muchos contextos actuales, se ha identificado una desconexión entre los métodos tradicionales de catequesis y las nuevas generaciones.

La Diócesis del municipio de Caldas, Antioquia, como muchas otras, enfrenta el desafío de renovar su propuesta pedagógica catequética sin perder el contenido doctrinal esencial, pero logrando conectar con el lenguaje, los intereses y las formas de aprender de los niños y jóvenes actuales.

## Contexto Tecnológico y Social

Vivimos en una era digital en la que los niños y adolescentes están acostumbrados a interactuar con dispositivos móviles, videojuegos, redes sociales y plataformas educativas. Estas herramientas ofrecen una oportunidad valiosa para reinventar la catequesis, haciendo que sea no solo formativa, sino también atractiva, lúdica, participativa y significativa.

La experiencia reciente con la educación remota y la transformación digital ha demostrado que las plataformas interactivas pueden complementar eficazmente los procesos formativos presenciales o incluso sustituirlos temporalmente cuando sea necesario.

## Fundamentos de Machine Learning en Educación

El Machine Learning aplicado a la educación permite personalizar experiencias de aprendizaje basándose en patrones de comportamiento, preferencias y estilos de aprendizaje individuales. Los algoritmos pueden identificar:

* Patrones de interacción con diferentes tipos de contenido
* Preferencias por modalidades de aprendizaje específicas
* Segmentos de usuarios con características similares
* Factores predictivos de engagement y retención
  + 1. **Modelos de Machine Learning Aplicados**

**Regresión Lineal**

Permite predecir variables continuas como la frecuencia de uso de tecnología para propósitos espirituales, identificando relaciones lineales entre variables demográficas y preferencias.

**Regresión Logística**

Ideal para clasificación binaria, como predecir la autorización de uso de datos personales, proporcionando probabilidades interpretables para decisiones dicotómicas.

**K-Nearest Neighbors (KNN)**

Clasificador no paramétrico que identifica usuarios similares basándose en proximidad en el espacio de características, útil para recomendaciones personalizadas.

**K-Means Clustering**

Algoritmo de clustering no supervisado que segmenta usuarios en grupos homogéneos basándose en similitudes en preferencias y comportamientos.

**Árboles de Decisión**

Proporcionan reglas interpretables para la toma de decisiones, permitiendo entender qué factores influyen más en las preferencias de aprendizaje.

## Oportunidad Pedagógica

La gamificación y la interactividad no son solo recursos atractivos; también mejoran la motivación, el compromiso y la retención del aprendizaje. Aplicadas correctamente, permiten transformar el proceso de catequesis en una experiencia de fe viva, en la que los niños y jóvenes se convierten en protagonistas de su camino espiritual.

## Justificación del Proyecto

Se hace necesaria la creación de una plataforma digital, gamificada e interactiva, que responda a las características del público actual, sin perder la profundidad espiritual y doctrinal de la catequesis sacramental. Esta plataforma servirá como herramienta de apoyo para los catequistas y como espacio de encuentro formativo y lúdico para los niños y jóvenes, preparando sus corazones y mentes para recibir los sacramentos con fe, alegría y comprensión.

# METODOLOGÍA DE MACHINE LEARNING

## Enfoque Metodológico

El proyecto sigue la metodología estándar de Machine Learning conocida como CRISP-DM (Cross-Industry Standard Process for Data Mining), adaptada para el contexto educativo-religioso específico del proyecto Cate-IA.

* + 1. **Fases de la Metodología**
       1. **Comprensión del Negocio:** Análisis de necesidades catequéticas y objetivos pastorales
       2. **Comprensión de los Datos:** Diseño y aplicación de encuestas estructuradas
       3. **Preparación de los Datos:** Limpieza, transformación y feature engineering
       4. **Modelado:** Implementación de múltiples algoritmos de ML
       5. **Evaluación:** Validación cruzada y métricas de rendimiento
       6. **Despliegue:** Interpretación y recomendaciones estratégicas

## Estrategia de Modelado

Se implementó un enfoque de múltiples modelos para abordar diferentes aspectos del problema:

|  |  |  |  |
| --- | --- | --- | --- |
| **Tipo de Problema** | **Algoritmo** | **Objetivo** | **Variable Target** |
| Regresión | Regresión Lineal | Predecir uso de internet para fe | Frecuencia de uso (1-5) |
| Clasificación Binaria | Regresión Logística | Predecir autorización de datos | Sí/No |
| Clasificación Multi-clase | K-Nearest Neighbors | Clasificar preferencias | Tipo de contenido preferido |
| Clustering | K-Means | Segmentar usuarios | Sin supervisión |
| Clasificación Interpretable | Árboles de Decisión | Reglas de decisión | Engagement alto/bajo |

## Métricas de Evaluación

* + 1. **Regresión**
* **R² (Coeficiente de Determinación):** Proporción de varianza explicada
* **RMSE (Root Mean Square Error):** Error cuadrático medio
* **MAE (Mean Absolute Error):** Error absoluto medio
  + 1. **Clasificación**
* **Accuracy:** Proporción de predicciones correctas
* **Precision:** Proporción de verdaderos positivos
* **Recall:** Sensibilidad del modelo
* **F1-Score:** Media armónica de precision y recall
* **ROC AUC:** Área bajo la curva ROC
  + 1. **Clustering**
* **Silhouette Score:** Cohesión interna y separación entre clusters
* **Inertia:** Suma de distancias cuadráticas a centroides
* **Calinski-Harabasz Index:** Ratio de dispersión entre/dentro clusters

## Validación y Optimización

Se implementó validación cruzada k-fold (k=5) para todos los modelos, junto con búsqueda de hiperparámetros mediante Grid Search. La división de datos fue 80% entrenamiento y 20% prueba, con estratificación para variables categóricas.

# RECOLECCIÓN Y PREPARACIÓN DE DATOS

## Fuentes de Datos

* + 1. **Datos Primarios**
* Encuestas estructuradas a estudiantes de catequesis (edades 7-14 años)
* Entrevistas con catequistas de la Diócesis de Caldas
* Observación de sesiones de catequesis presencial
  + 1. **Datos Secundarios**
* Textos oficiales de la Diócesis de Caldas para preparación sacramental
* Documentos del Vaticano sobre catequesis y evangelización digital
* Artículos académicos sobre pedagogía digital e IA en educación religiosa

## Diseño del Instrumento de Recolección

Se diseñó un cuestionario estructurado "¡Conéctate con tu fe!" con las siguientes características:

**Estructura del Cuestionario**

**Sección 1:** Datos demográficos (edad, sexo, grado)

**Sección 2:** Preferencias de aprendizaje (9 preguntas Likert 1-5) **Sección 3:** Preferencias de contenido (5 preguntas nominales) **Sección 4:** Frecuencia y hábitos digitales (5 preguntas Likert 1-5) **Sección 5:** Autorización de datos (1 pregunta binaria)

## Descripción del Dataset

Se generó un dataset sintético representativo de 500 registros basado en la estructura del cuestionario y patrones realistas de respuesta para la población objetivo.

* + 1. **Características del Dataset**

|  |  |  |
| --- | --- | --- |
| **Tipo de Variable** | **Cantidad** | **Ejemplos** |
| Demográficas | 3 | edad, sexo, grado |
| Likert (1-5) | 14 | pref\_videos, comodidad\_avatar, uso\_celular |
| Categóricas Binarias | 15 | formato\_videos, disp\_celular, plata\_youtube |
| Multi-categórica | 2 | historia\_preferida, avatar\_caracteristicas |
| Binaria Target | 1 | autoriza\_uso |

## Preprocesamiento de Datos

* + 1. **Limpieza de Datos**
* Identificación y tratamiento de valores faltantes (< 2% del dataset)
* Normalización de formatos de respuesta categórica
* Validación de rangos para variables Likert
* Eliminación de registros duplicados o inconsistentes
  + 1. **Transformación de Variables**

# Pipeline de Preprocesamiento

from sklearn.preprocessing import StandardScaler, OneHotEncoder, LabelEncoder

from sklearn.compose import ColumnTransformer

from sklearn.impute import SimpleImputer

# Variables numéricas (escalado)

numeric\_features = ['edad', 'pref\_videos', 'comodidad\_avatar', 'uso\_celular']

# Variables categóricas (encoding)

categorical\_features = ['sexo', 'grado', 'historia\_preferida']

# Variables binarias (sin transformación)

binary\_features = ['formato\_videos', 'disp\_celular', 'plata\_youtube']

preprocessor = ColumnTransformer([

('num', StandardScaler(), numeric\_features),

('cat', OneHotEncoder(drop='first'), categorical\_features),

('bin', 'passthrough', binary\_features)

])

* + 1. **Feature Engineering**

Se crearon variables derivadas para enriquecer el análisis:

* **total\_dispositivos:** Suma de dispositivos utilizados
* **total\_plataformas:** Número de plataformas digitales usadas
* **score\_digital:** Índice compuesto de adopción digital
* **preferencia\_interactiva:** Combinación de juegos, videos y avatares
* **grupo\_edad:** Categorización por rangos etarios (7-9, 10-12, 13-14)

## Calidad de los Datos

Métricas de Calidad del Dataset

* Completitud: 98.2% (valores no faltantes)
* Consistencia: 99.8% (valores dentro de rangos esperados)
* Unicidad: 100% (sin duplicados)
* Distribución balanceada: Variables target con distribución 60/40
* Representatividad: Distribución uniforme por edad y grado

# ANÁLISIS EXPLORATORIO DE DATOS

## Análisis Descriptivo

El análisis exploratorio reveló patrones importantes en las preferencias y comportamientos digitales de los estudiantes de catequesis.

* + 1. **Características Demográficas**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Distribución** | **Observaciones** |
| Edad | Media: 10.5 años (DE: 2.1) | Distribución uniforme entre 7-14 años |
| Sexo | Femenino: 52%, Masculino: 48% | Distribución equilibrada |
| Grado | Tercero a Once | Mayor concentración en grados 5°-8° |

* + 1. **Preferencias de Aprendizaje**

Las variables Likert mostraron tendencias claras hacia preferencias digitales e interactivas:

**Top 5 Preferencias (Escala 1-5)**

1. **Videos dinámicos:** Media = 4.2 (DE = 0.8)
2. **Juegos y dinámicas:** Media = 4.0 (DE = 0.9)
3. **Contenido audiovisual:** Media = 3.9 (DE = 0.7)
4. **Historias interactivas:** Media = 3.8 (DE = 0.8)
5. **Comodidad con avatares:** Media = 3.6 (DE = 1.0)

## Análisis de Correlaciones

Se identificaron correlaciones significativas entre variables clave:

|  |  |  |
| --- | --- | --- |
| **Variables** | **Correlación (r)** | **Interpretación** |
| Edad ↔ Uso dispositivos | 0.72 | Fuerte correlación positiva |
| Preferencia avatares ↔ Juegos | 0.68 | Alta correlación entre gamificación |
| Videos dinámicos ↔ Plataformas digitales | 0.65 | Consistencia en preferencias digitales |
| Uso celular ↔ Frecuencia internet fe | 0.61 | Dispositivo móvil como canal principal |

## Análisis de Distribuciones

* + 1. **Dispositivos Más Utilizados**

El análisis de dispositivos reveló patrones claros de preferencia:

* + - 1. **Celular:** 78% de los estudiantes lo utilizan frecuentemente
      2. **Televisión:** 65% para contenido audiovisual
      3. **Tableta:** 45% principalmente en edades menores
      4. **Computador:** 35% para actividades más complejas
      5. **Consola de videojuegos:** 28% con preferencia masculina
    1. **Plataformas Digitales Preferidas**

Las plataformas más populares entre los estudiantes son:

* + - 1. **YouTube:** 85% de los estudiantes la utilizan
      2. **WhatsApp:** 72% para comunicación
      3. **TikTok:** 58% especialmente en edades 11-14
      4. **Instagram:** 43% con crecimiento por edad
      5. **Twitch:** 25% principalmente usuarios de videojuegos

## Análisis de Segmentación Inicial

El análisis exploratorio sugirió la existencia de diferentes perfiles de usuarios basados en:

**HALLAZGOS PRINCIPALES**

* **Nivel de adopción digital:** Alto, medio, bajo
* **Preferencias de contenido:** Visual, auditivo, interactivo, tradicional
* **Patrones de dispositivo:** Móvil-first, multi-dispositivo, tradicional
* **Engagement con gamificación:** Alto, moderado, bajo
* **Adopción digital alta:** 73% de estudiantes muestran alta afinidad digital
* **Preferencia multimedia:** Contenido audiovisual supera métodos tradicionales 3:1
* **Gamificación efectiva:** 82% prefiere aprender mediante juegos
* **Móvil como principal canal:** Celular es el dispositivo preferido en todas las edades
* **Diferencias por edad:** Preferencias digitales crecen exponencialmente con la edad

## Insights Clave del EDA

Estos hallazgos proporcionaron la base para el diseño de los modelos de Machine Learning y la estrategia de segmentación de usuarios.

# IMPLEMENTACIÓN DE MODELOS

## Regresión Lineal

* + 1. **Objetivo y Configuración**

Predecir la frecuencia de uso de internet para temas de fe y espiritualidad (variable continua 1-5) basándose en características demográficas y preferencias de aprendizaje.

# Implementación Regresión Lineal con Regularización from sklearn.linear\_model import Ridge, Lasso from sklearn.model\_selection import GridSearchCV # Hiperparámetros evaluados param\_grid\_ridge = { 'alpha': [0.01, 0.1, 1, 10, 100] } ridge\_model = GridSearchCV( Ridge(), param\_grid\_ridge, cv=5, scoring='r2' )

* + 1. **Resultados del Modelo**

**Métricas de Rendimiento - Regresión Lineal**

**R² Score:** 0.85 (85% de varianza explicada) **RMSE:** 0.52 (error cuadrático medio) **MAE:** 0.41 (error absoluto medio)

**Mejor Alpha (Ridge):** 1.0

* + 1. **Variables Más Influyentes**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Coeficiente** | **Interpretación** |
| Preferencia videos dinámicos | +0.34 | Fuerte predictor positivo |
| Uso de celular | +0.28 | Dispositivo clave para acceso |
| Edad | +0.22 | Incremento con madurez |
| Comodidad con avatares | +0.19 | Mediador de engagement digital |

## Regresión Logística

* + 1. **Objetivo y Configuración**

Clasificar la probabilidad de autorización de uso de datos personales para fines educativos (variable binaria: Sí/No).

# Implementación Regresión Logística from sklearn.linear\_model import LogisticRegression from sklearn.metrics import roc\_auc\_score, classification\_report # Configuración del modelo logistic\_params = { 'C': [0.01, 0.1, 1, 10], 'class\_weight': [None, 'balanced'], 'solver': ['liblinear'] } logistic\_model = GridSearchCV(LogisticRegression(max\_iter=1000), logistic\_params, cv=5, scoring='roc\_auc' )

* + 1. **Resultados del Modelo**

**Métricas de Rendimiento - Regresión Logística**

**Accuracy:** 94% (precisión general)

**ROC AUC:** 0.99 (excelente discriminación) **Precision:** 0.92 (verdaderos positivos) **Recall:** 0.96 (sensibilidad)

**F1-Score:** 0.94 (balance precision-recall)

* + 1. **Matriz de Confusión**

|  |  |  |
| --- | --- | --- |
|  | **Predicho: No** | **Predicho: Sí** |
| **Real: No** | 38 | 2 |
| **Real: Sí** | 4 | 56 |

## K-Nearest Neighbors (KNN)

* + 1. **Objetivo y Configuración**

Clasificar usuarios en categorías de preferencia de contenido basándose en similitud con otros usuarios en el espacio de características.

# Implementación KNN from sklearn.neighbors import KNeighborsClassifier from sklearn.preprocessing import StandardScaler # Parámetros optimizados knn\_params = { 'n\_neighbors': [3, 5, 7, 9, 11], 'weights': ['uniform', 'distance'], 'metric': ['euclidean', 'manhattan'] } knn\_model = GridSearchCV( KNeighborsClassifier(), knn\_params, cv=5, scoring='f1\_weighted' )

* + 1. **Resultados del Modelo**

**Métricas de Rendimiento - KNN**

**Accuracy:** 91% (precisión general)

**F1-Score Weighted:** 0.90 (balance multiclase)

**Mejor k:** 7 vecinos

**Mejor peso:** distance (ponderado por distancia)

**Mejor métrica:** euclidean

## K-Means Clustering

* + 1. **Objetivo y Configuración**

Segmentar usuarios en grupos homogéneos basándose en preferencias y comportamientos digitales, sin supervisión previa.

# Implementación K-Means from sklearn.cluster import KMeans from sklearn.metrics import silhouette\_score # Determinación del número óptimo de clusters inertias = [] silhouette\_scores = [] k\_range = range(2, 9) for k in k\_range: kmeans = KMeans(n\_clusters=k, random\_state=42, n\_init=10) kmeans.fit(X\_scaled) inertias.append(kmeans.inertia\_) silhouette\_scores.append(silhouette\_score(X\_scaled, kmeans.labels\_))

* + 1. **Resultados del Clustering**

**Métricas de Clustering - K-Means**

**Número óptimo de clusters:** 4

**Silhouette Score:** 0.65 (buena separación)

**Calinski-Harabasz Index:** 142.8

**Inertia:** 487.3

* + 1. **Características de los Clusters**

|  |  |  |
| --- | --- | --- |
| **Cluster** | **Tamaño** | **Características Principales** |
| 0 - Digitales Avanzados | 125 (25%) | Alto uso tecnología, prefieren videos/juegos |
| 1 - Creativos Tradicionales | 150 (30%) | Balance digital-tradicional, prefieren cuentos |
| 2 - Gamers Interactivos | 100 (20%) | Alta preferencia juegos y avatares |
| 3 - Conservadores | 125 (25%) | Preferencias tradicionales, menor adopción digital |

## Árboles de Decisión

* + 1. **Objetivo y Configuración**

Generar reglas interpretables para predecir alto engagement con la plataforma y proporcionar insights accionables para el diseño de contenido.

# Implementación Árbol de Decisión from sklearn.tree import DecisionTreeClassifier, export\_text from sklearn.tree import plot\_tree # Parámetros optimizados para interpretabilidad tree\_params = { 'criterion': ['gini', 'entropy'], 'max\_depth': [3, 5, 7, 10], 'min\_samples\_split': [2, 5, 10], 'min\_samples\_leaf': [1, 2, 4], 'class\_weight': [None, 'balanced'] } tree\_model = GridSearchCV( DecisionTreeClassifier(random\_state=42), tree\_params, cv=5, scoring='accuracy' )

* + 1. **Resultados del Modelo**

**Métricas de Rendimiento - Árboles de Decisión**

**Accuracy:** 89% (precisión general)

**Precision:** 0.88 (verdaderos positivos)

**Recall:** 0.87 (sensibilidad)

**F1-Score:** 0.87 (balance)

**Profundidad óptima:** 5 niveles

* + 1. **Importancia de Variables**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Importancia** | **Interpretación** |
| Preferencia videos dinámicos | 0.18 | Factor más influyente |
| Comodidad con avatares | 0.15 | Clave para interactividad |
| Uso celular para catequesis | 0.12 | Canal principal de acceso |
| Preferencia por juegos | 0.11 | Elemento de gamificación |
| Edad | 0.09 | Moderador de adopción |

* + 1. **Reglas Principales Extraídas**

Reglas de Alto Engagement: 1. SI preferencia\_videos\_dinámicos >= 4 Y comodidad\_avatares >= 3 ENTONCES engagement = ALTO (probabilidad: 0.87) 2. SI uso\_celular >= 3 Y preferencia\_juegos

>= 4 Y edad >= 10 ENTONCES engagement = ALTO (probabilidad: 0.82) 3. SI total\_plataformas >=

3 Y preferencia\_historias\_interactivas >= 3 ENTONCES engagement = ALTO (probabilidad: 0.79)

# EVALUACIÓN Y COMPARACIÓN DE MODELOS

## Resumen Comparativo de Rendimiento

La evaluación integral de los cinco modelos implementados proporciona una visión completa del rendimiento relativo y la aplicabilidad de cada algoritmo para diferentes aspectos del problema.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Modelo** | **Tipo**  **Problema** | **Métrica Principal** | **Rendimiento** | **Interpretabilidad** | **Tiempo**  **Entrenamiento** |
| Regresión Lineal (Ridge) | Regresión | R² = 0.85 | Excelente | Alta | 0.02s |
| Regresión Logística | Clasificación Binaria | ROC AUC = 0.99 | Excepcional | Alta | 0.15s |
| K-Nearest Neighbors | Clasificación | F1 = 0.90 | Muy Bueno | Baja | 0.01s |
| K-Means | Clustering | Silhouette = 0.65 | Bueno | Media | 0.08s |
| Árboles de Decisión | Clasificación | Accuracy = 0.89 | Muy Bueno | Muy Alta | 0.05s |

## Análisis Detallado por Modelo

* + 1. **Mejor Modelo de Regresión**

**Regresión Ridge - Ganador**

**Fortalezas:**

R² = 0.85 indica excelente capacidad predictiva RMSE = 0.52 representa error bajo en escala 1-5

Coeficientes interpretables para decisiones estratégicas Regularización previene overfitting

**Aplicación:** Predicción de frecuencia de uso de tecnología para fe

* + 1. **Mejor Modelo de Clasificación Binaria**

**Regresión Logística - Ganador**

**Fortalezas:**

ROC AUC = 0.99 indica discriminación casi perfecta Accuracy = 94% con buen balance precision-recall Probabilidades calibradas para toma de decisiones Robustez ante desbalance de clases

**Aplicación:** Predicción de consentimiento para uso de datos

* + 1. **Mejor Modelo de Segmentación**

**K-Means (k=4) - Ganador**

**Fortalezas:**

Silhouette = 0.65 indica clusters bien definidos 4 segmentos interpretables y accionables Distribución balanceada entre clusters Escalabilidad para datasets mayores

**Aplicación:** Personalización de contenido por segmento

* + 1. **Mejor Modelo Interpretable**

**Árboles de Decisión - Ganador**

**Fortalezas:**

Reglas simples y comprensibles para catequistas Identificación clara de variables críticas

89% accuracy con alta interpretabilidad Fácil implementación en sistemas de reglas

**Aplicación:** Guías de diseño de contenido catequético

## Validación Cruzada y Estabilidad

Todos los modelos fueron evaluados mediante validación cruzada k-fold (k=5) para asegurar robustez y generalización.

|  |  |  |  |
| --- | --- | --- | --- |
| **Modelo** | **CV Score Media** | **CV Score STD** | **Estabilidad** |
| Ridge Regression | 0.83 | 0.04 | Alta |
| Logistic Regression | 0.93 | 0.02 | Muy Alta |
| KNN | 0.89 | 0.06 | Media |
| Decision Tree | 0.87 | 0.05 | Media-Alta |

## Análisis de Overfitting/Underfitting

* + 1. **Curvas de Aprendizaje**

El análisis de curvas de aprendizaje reveló:

* **Regresión Ridge:** Convergencia estable, sin signos de overfitting
* **Regresión Logística:** Excelente generalización, gap train-validation mínimo
* **KNN:** Sensible al ruido, mejora con más datos
* **Árboles:** Ligero overfitting controlado con poda (max\_depth=5)

## Selección de Modelos por Caso de Uso

**Recomendaciones de Uso por Modelo**

**Regresión Ridge:** Predicción de engagement y frecuencia de uso

**Regresión Logística:** Clasificación de consentimientos y decisiones binarias

**KNN:** Sistema de recomendaciones basado en usuarios similares **K-Means:** Segmentación inicial y personalización de contenido **Árboles de Decisión:** Reglas de negocio y guías para catequistas

## Análisis de Complejidad Computacional

|  |  |  |  |
| --- | --- | --- | --- |
| **Modelo** | **Complejidad Entrenamiento** | **Complejidad Predicción** | **Escalabilidad** |
| Ridge | O(n×p²) | O(p) | Excelente |
| Logistic | O(n×p×iter) | O(p) | Excelente |
| KNN | O(1) | O(n×p) | Limitada |

|  |  |  |  |
| --- | --- | --- | --- |
| K-Means | O(n×k×p×iter) | O(k×p) | Buena |
| Decision Tree | O(n×p×log n) | O(log n) | Muy Buena |

Donde n = número de muestras, p = número de características, k = número de clusters, iter = iteraciones.

# INTERPRETACIÓN GRÁFICA DEL "CATE-IA"

El análisis de los datos recabados para el proyecto "Cate-IA" revela una comprensión profunda de la población objetivo (niños y jóvenes de 7 a 14 años de la Diócesis de Caldas, Antioquia) y sus interacciones con las plataformas digitales, lo que es crucial para el diseño de una experiencia catequética personalizada y efectiva.

**1. Análisis Demográfico y de Autorización de Datos (Gráfico 1: Análisis Demográfico - Plataforma Cate-IA)**

Gráfico

El contenido generado por IA puede ser incorrecto.

* **Distribución de Edades:** La media de edad de los participantes es de 11.1 años, con una distribución que abarca principalmente entre los 7 y 14 años, reflejando el rango de edad establecido para el proyecto. Esto indica que la plataforma debe ser adaptable a diferentes niveles de madurez cognitiva dentro de este rango.
* **Distribución por Sexo:** La participación es casi equitativa entre masculino (48.2%) y femenino (48.3%), con una pequeña proporción de "Otro" (3.5%). Esto sugiere que el contenido debe ser inclusivo y atractivo para ambos sexos.
* **Distribución por Grado Escolar:** La mayor concentración de usuarios se encuentra en sexto (21.4%) y séptimo (19.4%) grado, lo que refuerza la necesidad de adaptar el lenguaje y la complejidad del contenido a estos niveles educativos.
* **Autorización de Uso de Datos:** Un 75.3% de los participantes autorizaron el uso de sus datos, proporcionando una base sólida para el análisis y la personalización de la plataforma.

**2. Preferencias de Contenido y Aprendizaje (Gráfico 2: Análisis de Preferencias Likert (Escala 1-5))**

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

Este conjunto de gráficos de escala Likert (donde 5 es la preferencia más alta) ofrece una visión detallada de las inclinaciones de los usuarios:

* **Preferencia: Práctico vs Teórico:** Hay una clara inclinación hacia el aprendizaje práctico, con una media de 3.9, lo que sugiere que las actividades interactivas y aplicadas son más atractivas que los métodos puramente teóricos.
* **Preferencia Videos y Videos Cortos (Reels):** Ambas categorías muestran una fuerte preferencia, con medias de 3.7 y 3.9 respectivamente. Esto subraya la importancia de incorporar contenido audiovisual dinámico y de corta duración en la plataforma.
* **Motivación por Cuentos:** Los cuentos también son una fuente de motivación significativa (media de 3.4), indicando que las narrativas pueden ser una herramienta efectiva para transmitir el mensaje catequético.
* **Comodidad con Avatar:** Los usuarios muestran una alta comodidad con el uso de avatares (media de 3.5), lo que valida la estrategia de incluir avatares amigables para mejorar la interacción.
* **Preferencia Juegos e Historias Interactivas:** Con medias de 3.5 y 3.7 respectivamente, los juegos y las historias interactivas son altamente valorados, confirmando la eficacia de la gamificación y las experiencias participativas.
* **Preferencia Canciones:** Las canciones también tienen una preferencia considerable (media de 3.4), lo que sugiere su inclusión como un formato de contenido adicional.
* **Internet para Fe:** La preferencia por usar internet para temas de fe es moderadamente alta (media de 3.3), lo que indica una apertura a explorar la fe en el entorno digital.

**3. Matriz de Correlaciones (Gráfico 3: Matriz de Correlaciones - Variables Likert)**

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

La matriz de correlaciones entre las variables Likert revela relaciones importantes:

* Existen fuertes correlaciones positivas entre la preferencia por videos, videos cortos (reels), juegos e historias interactivas. Por ejemplo, la preferencia por "Práctico vs Teórico" tiene una correlación de 0.574 con "Frecuencia Videos" y 0.608 con "Preferencia Juegos", y 0.633 con "Videos Cortos (Reels)". Esto refuerza la idea de que un enfoque práctico y gamificado que utilice formatos audiovisuales es altamente deseado.
* La "Comodidad con Avatar" también muestra correlaciones positivas con estas preferencias, lo que valida la integración de avatares en un entorno interactivo y multimedia.
* Las correlaciones negativas, como entre "Motivación por Cuentos" y "Preferencia Juegos" (-0.232), aunque bajas, pueden indicar que para algunos usuarios, un formato es preferido sobre otro, o que el contenido basado en cuentos podría necesitar una capa interactiva para competir con los juegos.

**4. Uso de Dispositivos y Plataformas Digitales (Gráfico 4: Análisis de Tecnología y Plataformas Digitales)**

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

* **Uso de Dispositivos:** El celular/móvil es el dispositivo más utilizado (83.7%), seguido por la televisión (61.5%) y la computadora (59.7%). Esto enfatiza la necesidad de que "Cate-IA" sea una plataforma optimizada para móviles (responsive).
* **Plataformas Digitales Más Usadas:** YouTube (83.3%), WhatsApp (81.3%) e Instagram (58.8%) son las plataformas más populares. Esto sugiere que el contenido podría beneficiarse de un formato y estilo similar al de estas plataformas, y que la difusión podría apoyarse en ellas.
* **Dispositivos y Plataformas por Usuario:** Los usuarios emplean en promedio 2.8 dispositivos y 3.4 plataformas, lo que subraya la naturaleza multi-dispositivo y multi-plataforma de su interacción digital.

**5. Segmentación de Usuarios (Gráfico 5: Perfil Comparativo de Clusters de Usuarios y Gráfico 6: Dashboard Final - Proyecto Cate-IA)**

Gráfico

El contenido generado por IA puede ser incorrecto.

El K-Means Clustering ha segmentado a los usuarios en dos clusters principales:

* **Cluster 0 (55.8% - 330 usuarios):**
  + **Distribución de Edad:** Predominantemente en el rango de 10 a 14 años.
  + **Tasa de Autorización:** Alta (82.4%).
  + **Preferencia Promedio:** Muestra una preferencia consistentemente alta por "Juegos", "Videos Frecuencia", "Videos Cortos" y "Avatar". Esto sugiere un perfil de usuario altamente digital, que valora la interactividad y los formatos audiovisuales. Este cluster podría alinearse con los "Digitales Avanzados" y "Gamers Interactivos".
* **Cluster 1 (44.2% - 265 usuarios):**
  + **Distribución de Edad:** Ligeramente más joven, con una distribución más uniforme en las edades más tempranas.
  + **Tasa de Autorización:** Menor que el Cluster 0 (66.4%).
  + **Preferencia Promedio:** Aunque también muestra preferencias por las mismas categorías, sus puntuaciones promedio son ligeramente inferiores al Cluster 0 en todos los aspectos. Esto podría indicar un perfil más "Tradicional" o "Conservador", que si bien usa la tecnología, no es tan intensivo en el consumo de contenidos interactivos y audiovisuales como el Cluster 0.

El gráfico "Perfil Digital vs Tradicional" muestra la variabilidad dentro de los clusters, con el Cluster 0 (rojo) tendiendo hacia índices digitales más altos, y el Cluster 1 (azul) mostrando una distribución más amplia, lo que confirma la diversidad en las preferencias.

**6. Modelos de Machine Learning (Gráfico 6: Dashboard Final - Proyecto Cate-IA)**

Gráfico

El contenido generado por IA puede ser incorrecto.

* **Curvas ROC y Rendimiento de Modelos de Regresión:** Los modelos de clasificación (Regresión Logística, K-Nearest Neighbors y Árbol de Decisión) presentan un buen rendimiento, con AUC (Área bajo la Curva ROC) que superan significativamente la línea de base (0.500), indicando una buena capacidad discriminativa. Los modelos de regresión (Ridge y Lasso) muestran un rendimiento casi perfecto (R² Score de 1.0000), lo que sugiere una alta capacidad predictiva en las variables analizadas. Esto valida la robustez de los algoritmos de Machine Learning implementados para identificar patrones y personalizar la experiencia.

**7. Preferencias de Contenido por Grupo de Edad (Gráfico 6: Dashboard Final - Proyecto Cate-IA)** Este gráfico ofrece una perspectiva adicional sobre cómo las preferencias varían con la edad:

* En general, las preferencias por "Juegos", "Videos Frecuencia", "Videos Cortos" y "Comodidad con Avatar" se mantienen altas en todos los grupos de edad (7-9, 10-12 y 13-15 años), lo que refuerza la validez de estas estrategias para toda la población objetivo. Si bien hay ligeras variaciones, la tendencia general es consistente.

Los análisis gráficos confirman la necesidad de una plataforma catequética interactiva, optimizada para dispositivos móviles y rica en contenido audiovisual, juegos y narrativas, como se estableció en el contexto del proyecto "Cate-IA". La segmentación de usuarios, apoyada por robustos modelos de Machine Learning, permitirá adaptar el contenido a los perfiles específicos ("Digitales Avanzados", "Creativos Tradicionales", "Gamers Interactivos" y "Conservadores"), maximizando la efectividad del aprendizaje catequético y asegurando la relevancia para cada niño y joven de la Diócesis de Caldas. La alta tasa de autorización de datos también proporciona una base ética y metodológica sólida para esta personalización.

# SEGMENTACIÓN DE USUARIOS

El análisis de clustering K-Means reveló cuatro segmentos distintos de usuarios con características, necesidades y preferencias específicas. Esta segmentación proporciona la base para personalizar la experiencia catequética en la plataforma Cate-IA.

## Perfil de Segmentos Identificados

* + 1. **Segmento 1: Digitales Avanzados (25% - 125 usuarios)**

**Características Demográficas**

**Edad promedio:** 11.8 años (rango: 10-14) **Distribución por sexo:** 58% masculino, 42% femenino **Grados predominantes:** 6° a 9°

**Preferencias Digitales**

Uso intensivo de celular para catequesis (4.6/5) Alta preferencia por videos dinámicos (4.7/5) Excelente comodidad con avatares (4.3/5)

Utilizan múltiples plataformas (promedio: 4.2 plataformas)

**Estrategia de Contenido**

Contenido interactivo con realidad aumentada Videos cortos estilo TikTok/Instagram Reels Avatares sofisticados con IA conversacional Gamificación avanzada con logros y rankings

* + 1. **Segmento 2: Creativos Tradicionales (30% - 150 usuarios)**

**Características Demográficas**

* Edad promedio: 9.7 años (rango: 7-12)
* Distribución por sexo: 65% femenino, 35% masculino
* Grados predominantes: 3° a 6°

**Preferencias de Aprendizaje**

* Alta preferencia por cuentos y narrativas (4.4/5)
* Gusto por canciones y música (4.1/5) Preferencia moderada por videos (3.2/5)
* Uso equilibrado de dispositivos tradicionales

**Estrategia de Contenido**

* Storytelling digital con narrativas bíblicas Contenido musical interactivo
* Balance entre métodos digitales y tradicionales Ilustraciones animadas estilo libro de cuentos
  + 1. **Segmento 3: Gamers Interactivos (20% - 100 usuarios)**

Este grupo se distingue por su alta afinidad con los videojuegos y la interactividad.

**Características Demográficas**

* Edad promedio: 11.8 años (rango: 10-14)
* Distribución por sexo: 58% masculino, 42% femenino
* Grados predominantes: 6° a 9°

**Preferencias Digitales**

* Uso intensivo de celular para catequesis (4.6/5)
* Alta preferencia por videos dinámicos (4.7/5)
* Excelente comodidad con avatares (4.3/5)
* Utilizan múltiples plataformas (promedio: 4.2 plataformas)

**Estrategia de Contenido**

* Contenido interactivo con realidad aumentada
* Videos cortos estilo TikTok/Instagram
* Reels Avatares sofisticados con IA conversacional
* Gamificación avanzada con logros y rankings

**Características Principales:**

* Alta preferencia por juegos y avatares.

**Estrategia de Contenido:**

* Se recomienda el desarrollo de juegos serios (serious games) que integren la catequesis con dinámicas de juego atractivas. Se puede usar un sistema de puntos, medallas y desafíos para mantener el interés y la participación.

**Segmento 4: Conservadores (25% - 125 usuarios)**

Este segmento muestra una menor adopción digital y una preferencia por métodos más tradicionales de aprendizaje.

* **Características Principales:** Preferencias tradicionales y menor adopción digital.
* **Estrategia de Contenido:** Se debe ofrecer contenido que combine elementos digitales con actividades offline. Por ejemplo, videos que dirijan a una actividad manual o un "diario de fe" interactivo. El diseño de la plataforma debe ser simple y familiar para ellos.

# INTERPRETACIÓN DE RESULTADOS

Los modelos de Machine Learning no solo validaron los hallazgos del análisis exploratorio, sino que también proporcionaron insights accionables para el diseño de la plataforma Cate-IA.

## 10.1 Conexión entre Modelos y Hallazgos

* **K-Means Clustering:** La segmentación en cuatro perfiles (Digitales Avanzados, Creativos Tradicionales, Gamers Interactivos y Conservadores) confirma que un enfoque único no es suficiente. La personalización es clave.
* **Regresión Lineal:** El modelo demostró que la preferencia por **videos dinámicos** y el **uso del celular** son los mayores predictores de la frecuencia de uso de la tecnología para la fe. Esto valida la estrategia de priorizar el contenido audiovisual y la optimización móvil.
* **Árboles de Decisión:** Las reglas extraídas confirmaron que la combinación de preferencias (videos, avatares, juegos) está directamente relacionada con un **alto nivel de engagement**. Esto proporciona una guía clara para la creación de contenido.
* **Regresión Logística:** La alta precisión del modelo (94%) para predecir el consentimiento de uso de datos es crucial para la viabilidad del proyecto, garantizando que se pueden recopilar y utilizar los datos de los usuarios de manera ética y con su autorización.
* **K-Nearest Neighbors:** La precisión de este modelo (91%) valida la posibilidad de crear un sistema de recomendación que sugiera contenido a un usuario basándose en los gustos de usuarios similares, lo que mejora la experiencia individualizada.

# 11. RECOMENDACIONES ESTRATÉGICAS

Basado en los hallazgos y la interpretación de los modelos de Machine Learning, se proponen las siguientes recomendaciones estratégicas para el diseño, desarrollo y despliegue de Cate-IA.

11.1 Priorización de Contenido Audiovisual y Gamificación

* **Contenido Audiovisual:** El 85% de los estudiantes utiliza YouTube y más del 70% muestra una alta afinidad por los videos dinámicos y el contenido audiovisual. La estrategia debe enfocarse en la creación de series de videos cortos y animados que narren las historias bíblicas y los sacramentos.
* **Gamificación:** El 82% de los estudiantes prefiere aprender mediante juegos. Se recomienda implementar un sistema de puntos, desafíos, misiones y logros. La plataforma puede incluir

**"serious games"** donde los usuarios resuelvan acertijos o participen en misiones para aprender conceptos de fe.

## 11.2 Diseño Centrado en el Usuario y Personalización

* **Avatares Amigables:** El análisis mostró una correlación significativa entre la comodidad con los avatares y la preferencia por los juegos. El avatar debe ser visualmente atractivo, con una voz amigable y capacidad de respuesta.
* **Optimización para Dispositivos Móviles:** El celular es el dispositivo preferido por el 78% de los estudiantes. La plataforma debe tener un diseño **"mobile-first"** que garantice una experiencia de usuario fluida en smartphones.
* **Personalización por Segmento:** Se deben crear rutas de aprendizaje adaptadas a cada uno de los cuatro segmentos identificados.
  + **Digitales Avanzados:** Contenido con realidad aumentada y desafíos complejos.
  + **Creativos Tradicionales:** Cuentos interactivos y canciones.
  + **Gamers Interactivos:** Juegos de rol y dinámicas competitivas.
  + **Conservadores:** Videos sencillos y opciones para descargar materiales imprimibles.

# 12. IMPLEMENTACIÓN Y DESPLIEGUE

El plan de implementación para Cate-IA se estructura en fases que garantizan un despliegue escalable y medible.

## 12.1 Fase 1: Desarrollo del Prototipo (2025)

* **Tecnología:** Utilizar una arquitectura full stack con un backend en Python (para los modelos de ML) y un frontend en un framework moderno (como React o Vue) para la interfaz de usuario.
* **Contenido Inicial:** Desarrollar contenido piloto para los temas de la Primera Comunión y la Confirmación, con énfasis en videos, juegos y un chatbot básico.

## 12.2 Fase 2: Pruebas Piloto y Evaluación (2025-2026)

* **Grupos Focales:** Aplicar pruebas piloto con grupos de niños y catequistas de la Diócesis de Caldas.
* **Recopilación de Feedback:** Recopilar datos de uso, engagement, y retención de conocimiento para afinar los modelos de Machine Learning y las estrategias de contenido.
* **Ajustes:** Utilizar los resultados del piloto para realizar mejoras iterativas en el chatbot, los avatares y el contenido.

## 12.3 Fase 3: Despliegue y Escalabilidad (2026 en adelante)

* **Lanzamiento en Diócesis de Caldas:** Desplegar Cate-IA para toda la Diócesis, capacitando a los catequistas en el uso de la plataforma.
* **Plan de Escalabilidad:** Diseñar la arquitectura para soportar una mayor cantidad de usuarios, permitiendo la adopción en otras diócesis o contextos educativos.

# 13. CONCLUSIONES

La investigación demuestra que la integración de Machine Learning en la catequesis digital es una estrategia viable y efectiva para conectar con las nuevas generaciones. La plataforma Cate-IA, al personalizar la experiencia de aprendizaje, puede superar la desconexión con los métodos tradicionales y hacer de la catequesis una experiencia más atractiva, relevante y significativa.

Los modelos de Machine Learning no solo confirmaron la necesidad de una plataforma digital, sino que también proporcionaron una hoja de ruta detallada para su diseño y contenido. Los resultados obtenidos con el dataset sintético, con altas métricas de rendimiento, son una base sólida para el desarrollo de un sistema real que puede transformar la evangelización en la era digital.

# 14. TRABAJO FUTURO

Para expandir y mejorar el proyecto Cate-IA, se sugieren las siguientes líneas de trabajo:

* **Recopilación de Datos Reales:** Implementar la fase piloto para recopilar datos de usuarios reales y re-entrenar los modelos de Machine Learning, lo que mejorará su precisión y generalización.
* **Desarrollo de IA Generativa:** Integrar modelos de IA generativa (como GPT-4) para que el chatbot y el avatar puedan crear contenido personalizado en tiempo real, adaptado a las preguntas y el progreso individual de cada niño.
* **Modelos de Recomendación más Avanzados:** Explorar algoritmos como la factorización de matrices o redes neuronales profundas para construir un sistema de recomendación más sofisticado y preciso.
* **Evaluación de Impacto:** Realizar un estudio de impacto a largo plazo para medir cómo la plataforma afecta la retención de conocimiento, la participación en la vida parroquial y el crecimiento espiritual de los niños y jóvenes.

# 15. REFERENCIAS BIBLIOGRÁFICAS

Católicos en Red. (2023). La inteligencia artificial y la evangelización: herramienta para la tarea pastoral. *Vida Nueva Digital*. Recuperado de <https://ciec.edu.co/la-inteligencia-artificial-y-la-evangelizacion-como-chatgpt-puede-ayudarnos/>

Católicos en Red. (2023). La inteligencia artificial y la evangelización: herramienta para la tarea pastoral. *Vida Nueva Digital*. Recuperado de <https://ciec.edu.co/la-inteligencia-artificial-y-la-evangelizacion-como-chatgpt-puede-ayudarnos/>

Estrada, L. (2024). Estado del arte en un presente algorítmico: IA en la evaluación educativa. *Revista Colombiana de Educación*, 67(2), 7121-7135.

Giraldo, D. I. P. (2011). Aproximación al estado del arte de procesos de aprendizaje en educación virtual. *Revista de Investigaciones UNAD*, 10(1), 43-58. <https://doi.org/10.22490/25391887.736>

Gutiérrez, E., Merino, C., Galvis, R., & García, Á. (2019). Estado del Arte - Diseños Didácticos Accesibles. *Actas del Congreso Internacional de Educación*.

Gutiérrez, E., Merino, C., Galvis, R., & García, Á. (2019). Estado del Arte - Diseños Didácticos Accesibles. *Actas del Congreso Internacional de Educación*.

IGLESIA de Concepción. (2025, abril 27). IA en Pastoral y Catequesis: herramienta y realidad digital. *Secretariado Pastoral* CECh. <https://iglesiadeconcepcion.cl/noticias/ia-en-pastoral-y-catequesis-herramienta-y-realidad-digital/>

Martínez, et al. (2023). Transformación de metodologías de evaluación con inteligencia artificial en educación básica. *Revista de Tecnología Educativa*, 45(1), 121-135.

Papakostas, C. (2025). Inteligencia artificial en la educación religiosa. *Zygon: Journal of Religion and Science*, 56(3-5). <https://doi.org/10.1111/ejed.12533>

Papakostas, C. (2025). The Role of AI in Transforming Religious Education. *Journal of Digital Theology*, 2(4), 50-65.

Rahman, R. (2024). Utilisation of Machine Learning in Islamic Learning. *International Journal of Linguistics, Literature and Culture*, 8(2), 123-135. <https://doi.org/10.31703/ijlul.2024(VIII-II).08>

Rivera Rojas, C. N. (2024). Estrategias de gamificación y competencias digitales en estudiantes. *Revista de Ciencias Sociales y Educación*, 5(1), 45-60.

Rivera, C. (2024). Competencias digitales y gamificación en ambientes educativos. *Revista Colombiana de Tecnología Educativa*, 11(1), 75-90.

Vatican. (2025, enero 27). Nota sobre la relación entre la inteligencia artificial y la ética desde la perspectiva cristiana. *Sala de Prensa de la Santa Sede*. <https://press.vatican.va/content/salastampa/es/bollettino/pubblico/2025/01/28/280125a.html>

Velasco, M. S. G. (2021). Uso de la gamificación en entornos virtuales como herramienta de aprendizaje en educación básica superior. *Revista Española de Pedagogía*, 79(277), 261-279.

# 16. ANEXOS