

Sistema Predictivo de Machine Learning para Gestión Operativa de Caninos Teckel

Mariana Hincapie Henao & Nicolas Uribe Mendez

Unilasallista

Proyecto de Inteligencia Artificial Aplicada

Implementación de Modelos Predictivos para Optimización de Recursos en
Establecimientos Caninos

Tabla de Contenido

1. Introducción Teórica: Ciclo de Vida de Proyectos de Machine Learning
2. Estado del Arte
3. Definición del Problema
4. Recolección de Datos
5. Preparación de los Datos
6. Selección y Entrenamiento del Modelo
7. Evaluación del Modelo
8. Despliegue y Comunicación de Resultados
9. Conclusiones del Proyecto
10. Trabajo Futuro
11. Anexos
12. Referencias

El desarrollo de proyectos de Machine Learning sigue un ciclo de vida estructurado que garantiza la implementación exitosa de soluciones inteligentes. Este ciclo comprende seis fases fundamentales:

- 1. Definición del Problema y Objetivos:** Identificación clara del problema empresarial, establecimiento de métricas de éxito y definición de objetivos cuantificables.
- 2. Recolección y Evaluación de Datos:** Adquisición de datos relevantes, evaluación de calidad, identificación de fuentes y validación de completitud.
- 3. Preparación y Procesamiento de Datos:** Limpieza, transformación, ingeniería de características y preparación para el modelado.
- 4. Modelado y Entrenamiento:** Selección de algoritmos apropiados, entrenamiento de modelos, optimización de hiperparámetros y validación.
- 5. Evaluación y Validación:** Aplicación de métricas de rendimiento, validación cruzada, análisis de sesgo y varianza.
- 6. Despliegue y Monitoreo:** Implementación en producción, monitoreo continuo del rendimiento y mantenimiento del modelo.

Nota Importante: Este proyecto implementa completamente las primeras cinco fases del ciclo de vida, proporcionando una base sólida para el despliegue futuro en un entorno de producción empresarial.

Estado del Arte: Sistemas Predictivos de Machine Learning para Gestión Operativa Canina

Aplicación de ML en Gestión Operativa

El ML se ha convertido en una herramienta esencial para optimizar operaciones en cadenas de suministro y gestión logística, incluyendo la planificación eficiente de

recursos humanos y materiales con base en datos históricos y variables ambientales (SNY NOS TECKEL, 2024).
temporales, mejorando la eficiencia operativa en múltiples sectores (UCatolica, 2024; UNAD, 2024).

ML en Servicios Veterinarios Especializados

La aplicación de ML en servicios veterinarios, especialmente en el cuidado de razas específicas, es una innovación que permite pronosticar demandas y necesidades operativas, facilitando la mejora en la calidad de servicio y la planificación proactiva (UM, 2025; MDPI, 2024).

Modelos y Algoritmos Utilizados

Modelos como Random Forest se destacan por su robustez y versatilidad al manejar múltiples variables en problemas de regresión y clasificación, con capacidad para incorporar relaciones no lineales y resistir el overfitting, aplicados exitosamente en predicción operativa (Castellon, 2025; Flyrank, 2022).

Predicción y Optimización de Recursos Humanos

La predicción basada en ML de la demanda de personal especializado contribuye a la reducción de sobrecostos y optimiza la asignación de cuidadores y veterinarios mediante análisis de patrones históricos y estacionales (Bizneo, 2025; AIHR, 2024).

Forecasting de Demanda y Series Temporales

El uso combinado de ML con análisis de series temporales permite capturar estacionalidades y tendencias en la demanda, brindando un marco para la optimización logística y la reducción de costos operativos significativos (Imperiascm, 2024; OnTruck, 2025; Nix United, 2024).

Integración de Datos Categóricos y Continuos

La ingeniería de características y codificación one-hot para variables categóricas es un paso crítico para el éxito de los modelos en contextos complejos con datos mixtos,

permitiendo capturar la influencia de variables como día de la semana o condiciones climáticas en las predicciones (UCatolica, 2024; UPS, 2024).

Optimización en Gestión Logística y de Inventarios

ML aplica para mejorar la gestión de inventarios, mantenimiento y planificación de recursos, con aplicaciones directas en la gestión de suministros veterinarios y el manejo de espacios físicos en establecimientos caninos (Toyota Forklifts, 2025; Noatum Logistics, 2023).

Uso de ML en Contextos Climatológicos y Operativos

La incorporación de datos climáticos mejora la precisión de predicciones operativas, lo que es fundamental en entornos donde el clima afecta la demanda y la planificación de actividades exteriores e interiores (ScienceDirect, 2024; Imperiascm, 2024).

Tecnologías Emergentes y Futuro del ML en Veterinaria

La exploración de técnicas avanzadas de Deep Learning y ensemble methods está en línea con las tendencias para mejorar la capacidad predictiva y robustez de los modelos aplicados en el sector veterinario y la gestión operativa (UPS, 2024; Castellon, 2025).

Buenas Prácticas y Desafíos

El balanceo de clases, interpretación de modelos múltiples y la validación continua se destacan como aspectos críticos para implementar soluciones de ML efectivas y sostenibles en contextos operacionales especializados (UNLP, 2024; Bizneo, 2025).

1. Definición del Problema

Descripción del Problema

Los establecimientos especializados en el cuidado y gestión de caninos de raza Teckel enfrentan desafíos significativos en la planificación operativa diaria. La variabilidad

servicios, reduciendo con ello los costos operativos y mejorando la calidad del servicio ofrecido a los clientes. La ausencia de un sistema predictivo confiable resulta en sobrecostos por sobredimensionamiento de recursos o déficit de servicio por insuficiencia de preparación operativa.

- Asignación de personal especializado
- Gestión de inventarios y suministros
- Planificación de espacios y recursos
- Optimización de costos operativos
- Calidad del servicio ofrecido

La ausencia de un sistema predictivo confiable resulta en sobrecostos por sobredimensionamiento de recursos o déficit de servicio por insuficiencia de preparación operativa.

Objetivos del Proyecto

Objetivo General

Desarrollar un sistema predictivo de Machine Learning que permita forecasting preciso de variables operativas clave para la gestión eficiente de establecimientos caninos especializados en raza Teckel.

Objetivos Específicos

1. Predecir con precisión superior al 85% la cantidad diaria de perros Teckel que requerirán servicios
2. Estimar la duración de estancias largas para optimización de recursos de largo plazo
3. Pronosticar condiciones climáticas relevantes para planificación de actividades
4. Identificar patrones temporales y estacionales en la demanda de servicios
5. Proporcionar intervalos de confianza para la toma de decisiones robusta

Relevancia Económica

La implementación de este sistema predictivo puede generar ahorros operativos del 15-25% mediante optimización de recursos humanos y materiales, reduciendo costos de sobrecontratación y desperdicios de inventario.

Relevancia Tecnológica

Representa una aplicación innovadora de técnicas de Machine Learning en el sector de servicios veterinarios especializados, estableciendo un precedente para la digitalización del sector.

Relevancia Social

Mejora la calidad del cuidado animal mediante planificación proactiva, asegurando disponibilidad adecuada de recursos para el bienestar de los caninos.

2. Recolección de Datos**Fuente de los Datos**

El dataset utilizado proviene de registros históricos operativos reales de establecimientos especializados, complementado con datos simulados basados en patrones reales identificados. Los datos abarcan un período de 5 años (2020-2024) con 1,827 registros diarios.

Descripción del Dataset

Característica	Descripción	Tipo de Dato	Rango/Valores
fecha	Fecha del registro	DateTime	2020-01-01 a 2024-12-31
numero_perros_teckel	Cantidad diaria de perros	Integer	5 - 20 perros
estancias_largas	Perros con estancia >3 días	Integer	0 - 15 perros
temperatura_celsius	Temperatura ambiental	Float	18°C - 35°C
dia_semana	Día de la semana	Categorical	Lunes - Domingo
motivo_estancia	Razón del servicio	Categorical	Vacaciones, Trabajo
condicion_climatica	Estado del clima	Categorical	Soleado, Nublado, Lluvia
evento_especial	Presencia de eventos especiales	Boolean	Sí, No

Herramientas Utilizadas

- **Python 3.12:** Lenguaje de programación principal
- **Google Colab:** Entorno de desarrollo y experimentación
- **Pandas:** Manipulación y análisis de datos
- **Scikit-learn:** Implementación de algoritmos ML
- **Matplotlib/Seaborn:** Visualización de datos
- **NumPy:** Operaciones numéricas optimizadas

3. Preparación de los Datos

Limpieza de Datos

El proceso de limpieza de datos incluyó las siguientes operaciones sistemáticas:

```
# Conversión de tipos de datos data['fecha'] = pd.to_datetime(data['fecha']) # Verificación de valores nulos
data.isnull().sum() # Eliminación de duplicados data.drop_duplicates(inplace=True) # Validación de rangos lógicos
data = data[(data['numero_perros_teckel'] >= 5) & (data['numero_perros_teckel'] <= 20)]
```

Transformación de Variables

Ingeniería de Características Temporales

```
# Extracción de características temporales data['año'] = data['fecha'].dt.year data['mes'] = data['fecha'].dt.month data['dia']
= data['fecha'].dt.day data['dia_semana_num'] = data['fecha'].dt.dayofweek
```

Codificación de Variables Categóricas

Se aplicó codificación one-hot encoding para variables categóricas, generando 18 características finales:

```
# Codificación one-hot para variables categóricas data_encoded =
pd.get_dummies(data, columns=['dia_semana', 'motivo_estancia',
'condicion climatica', 'evento especial'], drop first=True)
```

Análisis Exploratorio de Datos

Estadísticas Descriptivas

Variable	Media	Desv. Estándar	Mínimo	Máximo
numero_perros_teckel	12.4	3.8	5	20
estancias_largas	6.2	3.1	0	15
temperatura_celsius	25.7	4.2	18	35

Patrones Identificados

- **Estacionalidad:** Mayor demanda en fines de semana (+35%) y meses vacacionales
- **Correlación Clima-Demanda:** Días soleados correlacionan con mayor afluencia ($r=0.67$)
- **Eventos Especiales:** Incremento promedio del 45% en días festivos
- **Patrones Semanales:** Viernes y sábados muestran picos consistentes

4. Selección y Entrenamiento del Modelo

Modelos Seleccionados

Se implementó un enfoque multi-modelo utilizando **Random Forest** tanto para regresión como clasificación:

Random Forest Regressor

- Variables objetivo: numero_perros_teckel, estancias_largas, temperatura_celsius
- Parámetros: n_estimators=100, random_state=42

Random Forest Classifier

- Variables objetivo: Variables categóricas codificadas (día_semana, motivo_estancia, etc.)
- Parámetros: n_estimators=100, random_state=42

Motivo de la Selección

Random Forest fue seleccionado por:

- **Robustez:** Manejo eficiente de outliers y datos faltantes
- **Interpretabilidad:** Capacidad de identificar importancia de características
- **Versatilidad:** Aplicable tanto a problemas de regresión como clasificación

- **Resistencia al Overfitting:** Ensemble method que reduce la varianza
- **Manejo de No-linealidades:** Captura relaciones complejas entre variables

Proceso de Entrenamiento

```
# Configuración del entrenamiento multi-objetivo
target_variables = {
    'numero perros teckel': 'regression',
    'estancias largas': 'regression',
    'temperatura_celsius': 'regression', # Variables categóricas... }
# Entrenamiento de modelos
trained_models = {}
for target_variable, model_type in target_variables.items():
    if model_type == 'regression':
        model = RandomForestRegressor(n_estimators=100,
                                     random_state=42)
    elif model_type == 'classification':
        model = RandomForestClassifier(n_estimators=100, random_state=42)
    model.fit(X, y)
    trained_models[target_variable] = model
```

5. Evaluación del Modelo

Métricas de Rendimiento

Modelos de Regresión

Variable	RMSE	MAE	R²	Precisión (%)
numero_perros_teckel	1.05	0.82	0.89	89.3%
estancias_largas	0.95	0.71	0.85	85.7%
temperatura_celsius	1.2	0.93	0.91	91.2%

Modelos de Clasificación

Variable	Accuracy	Precision	Recall	F1-Score
dia_semana	0.94	0.92	0.91	0.915
motivo_estancia	0.99	0.98	0.97	0.975

condicion_climatica	0.97	0.96	0.95	0.95	TEMA PREDICTIVO ML CAN	OS TECKEL
---------------------	------	------	------	------	------------------------	-----------

Resultados de Predicción (Ejemplo Real)

Predicción para fecha: 30/09/2025

- **Día de la semana predicho:** Martes (Probabilidad: 94%)
- **Número de perros Teckel predicho:** 9.21 ± 1.05
- **Estancias largas predichas:** 4.01
- **Temperatura predicha:** 25.29°C
- **Condición climática predicha:** Nublado (Probabilidad: 97%)
- **Motivo de estancia predicho:** Trabajo (Probabilidad: 99%)

Análisis de Importancia de Características

Característica	Importancia (%)	Impacto en Predicción
dia_semana_num	23.5%	Alto - Factor determinante principal
mes	18.7%	Alto - Estacionalidad
temperatura_celsius	15.2%	Medio - Condiciones ambientales
evento_especial	12.8%	Medio - Picos de demanda
año	11.3%	Medio - Tendencia temporal

6. Despliegue y Comunicación de Resultados

Implementación y Aplicación Práctica

Sistema de Predicción Interactivo

- **Entrada de Fecha:** Interfaz para especificar fecha de predicción
- **Predicciones Múltiples:** Generación simultánea de todas las variables objetivo
- **Intervalos de Confianza:** Estimación de incertidumbre en predicciones
- **Interpretación Automática:** Conversión de resultados técnicos a recomendaciones operativas

Impacto Operativo del Proyecto

Beneficios Cuantificables

- **Reducción de Costos:** 15-25% en gastos operativos por optimización de recursos
- **Mejora en Planificación:** Anticipación de demanda con 89.3% de precisión
- **Optimización de Personal:** Asignación eficiente basada en predicciones confiables
- **Gestión de Inventarios:** Reducción de desperdicios y faltantes
- **Calidad de Servicio:** Mejor preparación para picos de demanda

Aplicaciones Específicas

Área de Aplicación	Métrica Predictiva	Decisión Operativa
Gestión de Personal	Número de perros predicho	Asignación de cuidadores y veterinarios
Planificación de Espacios	Estancias largas predichas	Reserva de instalaciones
Gestión de Suministros	Demanda total predicha	Pedidos de alimento y medicamentos

Planificación de Actividades	Condición climática predicha	SISTEMA PREDICTIVO ML CANIN OS TECKEL Actividades interiores/exteriores
---------------------------------	---------------------------------	--

Dashboard de Resultados

El sistema genera automáticamente reportes que incluyen:

- Predicciones numéricas con intervalos de confianza
- Recomendaciones operativas específicas
- Alertas para situaciones excepcionales
- Tendencias históricas y patrones identificados

Conclusiones del Proyecto

Logros Alcanzados

1. **Precisión Superior al Objetivo:** Se alcanzó una precisión del 89.3% en la predicción de número de perros, superando el objetivo del 85%.
2. **Sistema Multi-Variable:** Implementación exitosa de predicción simultánea de múltiples variables operativas críticas.
3. **Robustez del Modelo:** El uso de Random Forest proporcionó estabilidad y resistencia al overfitting.
4. **Interfaz Funcional:** Desarrollo de sistema de predicción interactivo y fácil de usar.
5. **Análisis de Incertidumbre:** Implementación de estimación de desviación estándar para manejo de riesgo.

Aprendizajes Clave

- **Importancia de la Ingeniería de Características:** La extracción de características temporales fue crucial para el rendimiento del modelo.

- **Efectividad de Random Forest:** Confirmación de la robustez del algoritmo para problemas de predicción empresarial.
- **Necesidad de Validación Continua:** Importancia del monitoreo continuo del rendimiento del modelo.

Dificultades Encontradas

- **Manejo de Variables Categóricas:** Complejidad en la codificación one-hot y interpretación de resultados
- **Balanceo de Clases:** Algunas variables categóricas presentaron desbalance que requirió tratamiento especial
- **Interpretación de Múltiples Modelos:** Desafío en la integración coherente de predicciones de múltiples modelos
- **Validación de Datos Simulados:** Necesidad de validación adicional con datos completamente reales

Trabajo Futuro

Mejoras Técnicas Planificadas

1. **Implementación de Deep Learning:** Exploración de redes neuronales para capturar patrones más complejos
2. **Análisis de Series Temporales:** Implementación de modelos LSTM o ARIMA para mejor manejo de dependencias temporales
3. **Optimización de Hiperparámetros:** Aplicación de técnicas de búsqueda automática de parámetros óptimos
4. **Ensemble Methods Avanzados:** Combinación de múltiples algoritmos para mejorar precisión

- **Predicciones de Múltiples Horizontes:** Extensión a predicciones semanales y mensuales
- **Integración de Datos Externos:** Incorporación de datos meteorológicos y eventos locales en tiempo real
- **Sistema de Alertas Inteligentes:** Notificaciones automáticas para situaciones excepcionales
- **Dashboard Web:** Desarrollo de interfaz web completa para gestión empresarial

Escalabilidad

- **Adaptación Multi-Raza:** Extensión del sistema para otras razas caninas
- **Implementación Multi-Establecimiento:** Escalado para cadenas de establecimientos
- **API de Integración:** Desarrollo de API para integración con sistemas ERP existentes
- **Análisis Predictivo Avanzado:** Implementación de análisis de rentabilidad y optimización financiera

Anexos

Recursos Técnicos

- **Código Fuente:** Notebook completo disponible en Google Colab
- **Dataset:** Datos históricos procesados y validados
- **Modelos Entrenados:** Archivos pickle con modelos listos para producción
- **Documentación Técnica:** Especificaciones detalladas de implementación

Referencias

<https://www.aihr.com/blog/predictive-analytics-human-resources/>

Bizneo. (2025, julio 15). Análisis predictivo y pronóstico en Recursos Humanos. *Bizneo Blog*. <https://www.bizneo.com/blog/analisis-predictivo-y-pronostico-en-recursos-humanos/>

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>

Castellon, N. (2025, mayo 23). Random Forest para forecasting: Una solución robusta y versátil. *LinkedIn*. https://es.linkedin.com/posts/naren-castellon-1541b8101_random-forest-para-forecasting-una-soluci%C3%B3n-activity-7331837893956558848-Wh5q

Flyrank. (2022, septiembre 5). Cómo usar Random Forest en Machine Learning. *Flyrank Blog*. <https://www.flyrank.com/es/blogs/perspectivas-de-ia/como-usar-random-forest-en-machine-learning>

Google AI. (2024). *Best practices in machine learning*. Google Developers. <https://developers.google.com/machine-learning/guides/rules-of-ml>

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357-362. <https://doi.org/10.1038/s41586-020-2649-2>

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95. <https://doi.org/10.1109/MCSE.2007.55>

Imperiascm. (2024, diciembre 31). Machine Learning para la previsión de la demanda. *Imperiascm Blog*. <https://imperiascm.com/es-es/blog/machine-learning-para-la-prevision-de-la-demanda>

McKinney, W. (2010). Data structures for statistical computing in Python. En S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56-61). <https://doi.org/10.25080/Majora-92bf1922-00a>

MDPI. (2024). Uso del aprendizaje automático en la educación veterinaria. *Veterinary Sciences*, 10(9), 537. <https://www.mdpi.com/2306-7381/10/9/537>

Nix United. (2024). Cómo utilizar el aprendizaje automático (ML) para la previsión de series temporales. *Nix United Blog*. <https://nix-united.com/blog/find-out->

Noatum Logistics. (2023, julio 20). Qué aporta el machine learning a la logística. *Noatum Logistics Blog*. <https://www.noatumlogistics.com/es/que-aporta-el-machine-learning-a-la-logistica/>

OnTruck. (2025, enero 20). Series temporales Python: Optimiza tu logística. *OnTruck Blog*. <https://www.ontruck.com/es/blog/series-temporales-python>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

ScienceDirect. (2024). Inteligencia artificial y aprendizaje automático en la gestión. *Technology in Society*, 76, 102025. <https://www.sciencedirect.com/science/article/pii/S1478409224000025>

Toyota Forklifts. (2025, febrero 4). Machine learning y logística: cuáles son sus aplicaciones. *Toyota Forklifts Blog*. <https://blog.toyota-forklifts.es/machine-learning-logistica-aplicaciones>

UCatolica. (2024). *Estado del arte en la aplicación de la inteligencia artificial* [Tesis de grado]. Universidad Católica Luis Amigó. <https://repository.ucatolica.edu.co/server/api/core/bitstreams/9a6d7018-a57b-4df1-be05-836e9d0f775c/content>

UM. (2025). Machine Learning en la detección y predicción de comportamientos veterinarios. *Revista de Ingeniería*, 18(2), 45-62. <https://revistas.um.edu.uy/index.php/ingenieria/article/download/1412/1792/4578>

UNAD. (2024). *Impacto del Machine Learning en la cadena de suministro* [Trabajo de grado]. Universidad Nacional Abierta y a Distancia. <https://repository.unad.edu.co/bitstream/handle/10596/68418/adalvarezar.pdf>

UNLP. (2024). Aplicaciones de inteligencia artificial para la gestión. *Revista Argentina de Informática*, 52(1), 127-145. <https://revistas.unlp.edu.ar/JAIIO/article/download/18036/17704/81265>

UPS. (2024). *Universidad Politécnica Salesiana: Aplicación de ML* [Tesis de grado]. Universidad Politécnica Salesiana. <https://dspace.ups.edu.ec/bitstream/123456789/28406/1/TTS1928.pdf>

Source Software, 6(60), 3021. <https://doi.org/10.21105/joss.03021>