

PROPUESTA DE TESIS DOCTORAL

**ESCUELA DOCTORAL FRANCO-PERUANA EN CIENCIAS DE LA
VIDA**

Convocatoria 2025

Factores Determinantes de la Sostenibilidad y Productividad Agropecuaria en el Perú: Análisis Predictivo mediante Big Data para el Desarrollo de Políticas Públicas de Desarrollo Rural Sostenible

Postulante:

FRED TORRES CRUZ

DNI: []

fred.torres@[email].com

Tel: +51 920 101 015

Puno, Perú

Agosto 2025

Índice

1. INFORMACIÓN GENERAL	3
1.1. Título de la Propuesta	3
1.2. Nombre del Postulante	3
1.3. Co-directores de Tesis	3
1.3.1. Co-director Francés	3
1.3.2. Co-director Peruano (UPCH)	3
2. RESUMEN	3
3. INTRODUCCIÓN/MARCO TEÓRICO	4
3.1. Contexto Nacional del Sector Agropecuario Peruano	4
3.2. Fundamentos Teóricos	5
3.2.1. Teoría de Sistemas Socio-Ecológicos Complejos	5
3.2.2. Paradigma del Big Data en Agricultura de Precisión	5
3.2.3. Desarrollo Rural Basado en Evidencia	5
3.3. Estado del Arte y Brechas de Conocimiento	6
4. PROBLEMÁTICA CIENTÍFICA Y PREGUNTA DE INVESTIGACIÓN	6
4.1. Problemática Central	6
4.2. Pregunta Principal de Investigación	7
4.3. Preguntas Específicas	7
5. HIPÓTESIS/OBJETIVOS DEL PROYECTO DE TESIS	7
5.1. Hipótesis	7
5.2. Objetivo General	7
5.3. Objetivos Específicos	8
6. BREVE DESCRIPCIÓN DE MATERIALES Y MÉTODOS	8
6.1. Fuente de Datos	8
6.2. Metodología	8
6.2.1. Fase 1: Preprocesamiento y Exploración (Meses 1-6)	8
6.2.2. Fase 2: Modelamiento con Machine Learning (Meses 7-18)	9
6.2.3. Fase 3: Desarrollo de Herramientas (Meses 19-30)	9
6.2.4. Fase 4: Validación y Transferencia (Meses 31-36)	9
6.3. Software y Herramientas	9
7. RESULTADOS PRELIMINARES	9
7.1. Calidad de Datos	10
7.2. Hallazgos Preliminares	10

8. CRONOGRAMA DE EJECUCIÓN	10
8.1. Año 1 - UPCH (Perú)	10
8.2. Año 2 - Universidad Francesa/IRD	11
8.3. Año 3 - UPCH (Perú)	11
9. REFERENCIAS BIBLIOGRÁFICAS RELEVANTES	11

1. INFORMACIÓN GENERAL

1.1. Título de la Propuesta

**Factores Determinantes de la Sostenibilidad y Productividad Agropecuaria en el Perú:
Análisis Predictivo mediante Big Data para el Desarrollo de Políticas Públicas de Desarrollo Rural Sostenible**

1.2. Nombre del Postulante

FRED TORRES CRUZ

1.3. Co-directores de Tesis

1.3.1. Co-director Francés

- **Nombre:** Dr. []
- **Afiliación:** []
- **Unidad Mixta de Investigación:** []
- **Teléfono:** []
- **Correo electrónico:** []

1.3.2. Co-director Peruano (UPCH)

- **Nombre:** Dr. []
- **Afiliación:** Universidad Peruana Cayetano Heredia
- **Facultad:** [Facultad de Ciencias y Filosofía / Facultad de Ciencias e Ingeniería]
- **Teléfono:** []
- **Correo electrónico:** []@upch.pe

2. RESUMEN

La presente investigación doctoral propone desarrollar un modelo predictivo integral basado en técnicas avanzadas de *big data* y *machine learning* para identificar y cuantificar los factores determinantes de la sostenibilidad y productividad en los sistemas agropecuarios peruanos.

El estudio se fundamenta en el análisis exhaustivo de una base de datos única que comprende 1,679 variables detalladas de productores agropecuarios distribuidos en las 25 regiones del Perú,

representando la diversidad agroecológica nacional. Esta riqueza de información permite una caracterización sin precedentes del sector agropecuario peruano.

La metodología combina técnicas de análisis estadístico multivariado, algoritmos de aprendizaje automático (Random Forest, Support Vector Machines, Redes Neuronales Profundas, XGBoost) y análisis espacial georreferenciado. Esta aproximación interdisciplinaria permitirá identificar patrones complejos no detectables mediante métodos tradicionales y desarrollar herramientas predictivas robustas.

Los resultados esperados incluyen: (i) la identificación precisa de factores críticos de éxito diferenciados por región y sistema productivo, (ii) el desarrollo de modelos predictivos con capacidad de estimar impactos ex-ante de intervenciones, (iii) la generación de un sistema de soporte de decisiones basado en evidencia, y (iv) recomendaciones de política pública operacionalizables y territorialmente diferenciadas.

El impacto potencial es significativo: la implementación de las recomendaciones podría incrementar la productividad agropecuaria nacional en 20-25 % mientras se conservan los recursos naturales, beneficiando directamente a los 2.3 millones de productores agropecuarios del país y contribuyendo a los Objetivos de Desarrollo Sostenible 2030.

3. INTRODUCCIÓN/MARCO TEÓRICO

3.1. Contexto Nacional del Sector Agropecuario Peruano

El Perú enfrenta el desafío crítico de transformar su sector agropecuario para garantizar la seguridad alimentaria de una población creciente mientras preserva su excepcional biodiversidad y recursos naturales. El sector agropecuario representa el 5.2 % del PBI nacional y emplea al 25 % de la población económicamente activa rural, constituyendo la principal fuente de sustento para millones de familias peruanas.

Sin embargo, el sector enfrenta desafíos estructurales significativos:

1. **Baja productividad sistémica:** El 70 % de los productores son pequeños agricultores con rendimientos 40 % inferiores al promedio latinoamericano, evidenciando brechas tecnológicas y de conocimiento críticas.
2. **Degradación ambiental acelerada:** El 38 % de los suelos agrícolas presenta algún grado de degradación, comprometiendo la capacidad productiva futura y la resiliencia del sistema alimentario.
3. **Alta vulnerabilidad climática:** Los eventos extremos asociados al cambio climático generan pérdidas anuales equivalentes al 15 % de la producción nacional, afectando desproporcionadamente a pequeños productores.

4. **Limitada adopción tecnológica:** Solo el 23 % de productores utiliza prácticas agrícolas mejoradas, revelando barreras estructurales para la innovación y transferencia tecnológica.
5. **Fragmentación y desarticulación:** La atomización de la propiedad (promedio 3.3 ha) y la débil articulación con mercados limitan las economías de escala y el acceso a servicios.

3.2. Fundamentos Teóricos

3.2.1. Teoría de Sistemas Socio-Ecológicos Complejos

La investigación adopta el marco conceptual de sistemas socio-ecológicos complejos (Ostrom, 2009), reconociendo que los sistemas agropecuarios son sistemas adaptativos caracterizados por:

- **Interdependencias múltiples:** Interacciones no lineales entre componentes biofísicos, económicos, sociales e institucionales que generan propiedades emergentes.
- **Retroalimentaciones dinámicas:** Bucles de retroalimentación positiva y negativa que pueden amplificar o amortiguar cambios en el sistema.
- **Umbrales críticos y resiliencia:** Puntos de inflexión que pueden generar transiciones abruptas entre estados del sistema, determinando su capacidad adaptativa.

3.2.2. Paradigma del Big Data en Agricultura de Precisión

Siguiendo a Wolfert et al. (2017), el paradigma del *big data* en agricultura se caracteriza por las "5 V": Volumen (grandes cantidades de datos), Velocidad (procesamiento en tiempo real), Variedad (múltiples fuentes), Veracidad (calidad de información) y Valor (generación de insights accionables).

La aplicación de técnicas de *machine learning* permite:

- Identificación de patrones complejos no lineales
- Predicción de resultados con alta precisión
- Optimización de recursos y decisiones
- Personalización de intervenciones

3.2.3. Desarrollo Rural Basado en Evidencia

El enfoque de políticas basadas en evidencia (OECD, 2020) enfatiza la importancia de utilizar datos rigurosos y análisis científico para el diseño de intervenciones públicas efectivas, maximizando el retorno social de la inversión.

3.3. Estado del Arte y Brechas de Conocimiento

La revisión sistemática de literatura revela avances internacionales significativos pero brechas críticas en el contexto peruano:

Avances internacionales relevantes:

- Zhang et al. (2021) desarrollaron modelos de *deep learning* para optimización de rendimientos en China, logrando mejoras del 18 % en precisión predictiva.
- Kumar et al. (2020) aplicaron Random Forest para identificar determinantes de adopción tecnológica en India.
- Pretty et al. (2018) establecieron marcos para intensificación sostenible, pero con limitada validación empírica en países megadiversos.

Brechas identificadas:

1. **Brecha metodológica:** Ausencia de estudios que integren big data con enfoque de sistemas complejos en contextos de alta diversidad agroecológica.
2. **Brecha empírica:** Falta de evidencia sobre factores determinantes específicos al contexto peruano.
3. **Brecha de aplicación:** Desconexión entre hallazgos científicos y diseño de políticas operacionalizables.

4. PROBLEMÁTICA CIENTÍFICA Y PREGUNTA DE INVESTIGACIÓN

4.1. Problemática Central

El problema central es la **ausencia de un entendimiento científico integral y cuantitativo de los factores multidimensionales que determinan la sostenibilidad y productividad de los sistemas agropecuarios peruanos**, lo cual limita severamente la capacidad del Estado para diseñar intervenciones efectivas de desarrollo rural.

Esta problemática se manifiesta en:

1. **Dimensión epistemológica:** Predominio de aproximaciones reduccionistas que analizan factores aislados sin considerar interacciones sistémicas.
2. **Dimensión metodológica:** Subutilización de tecnologías de análisis de datos masivos, limitando la extracción de conocimiento actionable.
3. **Dimensión práctica:** Desarticulación entre generación de conocimiento y toma de decisiones, resultando en políticas ineficientes.

4.2. Pregunta Principal de Investigación

¿Cuáles son los factores determinantes multidimensionales de la sostenibilidad y productividad en los sistemas agropecuarios peruanos, y cómo pueden ser modelados mediante técnicas de big data para desarrollar herramientas predictivas que orienten políticas públicas diferenciadas de desarrollo rural sostenible?

4.3. Preguntas Específicas

1. ¿Qué tipologías de productores pueden identificarse mediante clustering multivariado y cuáles son sus características distintivas?
2. ¿Cuáles variables técnicas, económicas, sociales y ambientales predicen mejor el desempeño sostenible?
3. ¿Cómo varían los factores determinantes según región agroecológica y sistema productivo?
4. ¿Qué modelos de machine learning presentan mejor capacidad predictiva?
5. ¿Cómo traducir los hallazgos en recomendaciones operacionalizables de política pública?

5. HIPÓTESIS/OBJETIVOS DEL PROYECTO DE TESIS

5.1. Hipótesis

La sostenibilidad y productividad de los sistemas agropecuarios peruanos están determinadas por una configuración específica y cuantificable de factores multidimensionales - técnicos (tecnología, prácticas agrícolas), económicos (acceso a mercados, crédito), sociales (capital social, educación), ambientales (calidad de suelos, agua) e institucionales (gobernanza) - cuyas interacciones complejas pueden ser modeladas mediante algoritmos de machine learning para desarrollar herramientas predictivas que permitan el diseño de políticas públicas diferenciadas y basadas en evidencia.

5.2. Objetivo General

Desarrollar un marco analítico integral basado en técnicas de big data y machine learning para identificar, cuantificar y modelar los factores determinantes de la sostenibilidad y productividad en los sistemas agropecuarios peruanos, generando herramientas predictivas y recomendaciones de política pública que contribuyan al desarrollo rural sostenible.

5.3. Objetivos Específicos

1. Caracterizar y tipificar los sistemas agropecuarios peruanos mediante técnicas de clustering multivariado, identificando patrones de sostenibilidad y productividad.
2. Identificar y jerarquizar los factores críticos determinantes del desempeño sostenible mediante algoritmos de machine learning (Random Forest, XGBoost, redes neuronales).
3. Desarrollar modelos predictivos robustos para identificar productores con alto potencial de mejora y estimar impactos ex-ante de intervenciones.
4. Generar mapas de vulnerabilidad y oportunidad mediante análisis espacial avanzado.
5. Formular recomendaciones de política pública diferenciadas por territorio y sistema productivo.

6. BREVE DESCRIPCIÓN DE MATERIALES Y MÉTODOS

6.1. Fuente de Datos

La investigación utilizará una base de datos excepcional con las siguientes características:

Tabla 1: Características de la Base de Datos Principal

Característica	Descripción
Cobertura geográfica	Nacional (25 regiones)
Número de variables	1,679 variables estructuradas
Dimensiones evaluadas	Producción, tecnología, economía, ambiente, social, institucional
Georreferenciación	Coordenadas GPS por unidad productiva
Representatividad	Factor de expansión nacional

6.2. Metodología

6.2.1. Fase 1: Preprocesamiento y Exploración (Meses 1-6)

- Auditoría de calidad de datos y tratamiento de valores faltantes
- Creación de índices compuestos de sostenibilidad
- Análisis exploratorio y visualización inicial
- Reducción dimensional (PCA, t-SNE)

6.2.2. Fase 2: Modelamiento con Machine Learning (Meses 7-18)

- **Clustering:** K-means, DBSCAN, Gaussian Mixture Models
- **Clasificación/Regresión:** Random Forest, XGBoost, SVM, Redes Neuronales
- **Análisis espacial:** Autocorrelación espacial, Kriging, GWR
- **Validación:** Cross-validation, bootstrap, métricas de desempeño

6.2.3. Fase 3: Desarrollo de Herramientas (Meses 19-30)

- Modelos ensemble y meta-learning
- Dashboard interactivo con visualizaciones dinámicas
- Simulación de escenarios de política
- Sistema de recomendaciones automatizado

6.2.4. Fase 4: Validación y Transferencia (Meses 31-36)

- Validación con panel de expertos (método Delphi)
- Pruebas piloto en territorios seleccionados
- Desarrollo de manuales y guías de política
- Capacitación a funcionarios públicos

6.3. Software y Herramientas

- **Lenguajes:** Python 3.9+, R 4.1+
- **Machine Learning:** scikit-learn, TensorFlow, XGBoost
- **Análisis espacial:** QGIS, GeoPandas, Google Earth Engine
- **Visualización:** Plotly, Dash, Tableau

7. RESULTADOS PRELIMINARES

Se ha realizado un análisis exploratorio inicial que confirma la viabilidad de la investigación:

7.1. Calidad de Datos

- Completitud: 92 % de datos completos en variables críticas
- Consistencia: 95 % de registros pasan validaciones lógicas
- Representatividad: Cobertura adecuada de todas las regiones agroecológicas

7.2. Hallazgos Preliminares

1. **Heterogeneidad productiva:** Coeficiente de variación ≥ 0.8 en productividad, confirmando la necesidad de enfoques diferenciados.
2. **Patrones espaciales:** Autocorrelación espacial significativa (Moran's I = 0.34, $p < 0.001$), indicando clustering geográfico de productividad.
3. **Factores tecnológicos:** Correlación positiva entre adopción tecnológica y productividad ($r = 0.42$), con efecto multiplicativo de tecnologías combinadas.
4. **Validación técnica:** Modelos preliminares de Random Forest alcanzan $R^2 \geq 0.75$ en predicción de productividad.

8. CRONOGRAMA DE EJECUCIÓN

8.1. Año 1 - UPCH (Perú)

Tabla 2: Cronograma detallado - Año 1

Período	Actividades	Productos
Meses 1-3	<ul style="list-style-type: none"> • Revisión sistemática de literatura • Limpieza y validación de datos • Cursos metodológicos UPCH 	<ul style="list-style-type: none"> • Estado del arte • Base datos depurada • Plan metodológico
Meses 4-6	<ul style="list-style-type: none"> • Análisis exploratorio • Creación índices sostenibilidad • Mapeo geográfico inicial 	<ul style="list-style-type: none"> • Reporte descriptivo • Índices validados • Atlas preliminar
Meses 7-9	<ul style="list-style-type: none"> • Análisis clustering • Primeros modelos ML • Redacción artículo 1 	<ul style="list-style-type: none"> • Tipologías definidas • Modelos baseline • Artículo enviado
Meses 10-12	<ul style="list-style-type: none"> • Análisis econométrico • Preparación Francia • Reporte anual 	<ul style="list-style-type: none"> • Modelos validados • Plan trabajo Francia • Informe IRD-UPCH

8.2. Año 2 - Universidad Francesa/IRD

Laboratorio: [A definir con co-director francés]

Supervisor: [A confirmar]

Tabla 3: Cronograma detallado - Año 2

Período	Actividades	Productos
Meses 13-15	<ul style="list-style-type: none"> • Integración laboratorio • ML avanzado • Colaboración internacional 	<ul style="list-style-type: none"> • Modelos refinados • Red colaboración • Certificaciones
Meses 16-18	<ul style="list-style-type: none"> • Análisis espacial • Deep learning • Artículo 2 	<ul style="list-style-type: none"> • Mapas detallados • Modelos DL • Artículo en revisión
Meses 19-21	<ul style="list-style-type: none"> • Validación cruzada • Análisis comparativo • Desarrollo software 	<ul style="list-style-type: none"> • Modelos validados • Estudio comparativo • Software v1.0
Meses 22-24	<ul style="list-style-type: none"> • Dashboard interactivo • Simulación políticas • Preparación retorno 	<ul style="list-style-type: none"> • Interfaz funcional • Escenarios simulados • Plan aplicación Perú

8.3. Año 3 - UPOCH (Perú)

Tabla 4: Cronograma detallado - Año 3

Período	Actividades	Productos
Meses 25-27	<ul style="list-style-type: none"> • Traducción en políticas • Validación expertos • Talleres stakeholders 	<ul style="list-style-type: none"> • Propuestas política • Validación nacional • Manual políticas
Meses 28-30	<ul style="list-style-type: none"> • Piloto implementación • Capacitación funcionarios • Análisis costo-beneficio 	<ul style="list-style-type: none"> • Piloto 3 regiones • 50 capacitados • Evaluación económica
Meses 31-33	<ul style="list-style-type: none"> • Redacción tesis • Artículos finales • Conferencias 	<ul style="list-style-type: none"> • Tesis completa • 2 artículos más • 3 presentaciones
Meses 34-36	<ul style="list-style-type: none"> • Pre-sustentación • Sustentación final • Disseminación 	<ul style="list-style-type: none"> • Aprobación comité • Grado Doctor • Plan postdoctoral

9. REFERENCIAS BIBLIOGRÁFICAS RELEVANTES

1. Barrett, C. B., Christiaensen, L., Sheahan, M., & Shimeles, A. (2017). On the structural transformation of rural Africa. *Journal of African Economies*, 26(suppl_1), i11-i35.

2. FAO. (2021). *The State of Food and Agriculture 2021: Making agrifood systems more resilient to shocks and stresses*. Rome: FAO.
3. Kumar, A., Padhee, A. K., & Kumar, S. (2020). How Indian agriculture should change to feed a growing population. *Food Security*, 12(2), 237-253.
4. Lipper, L., Thornton, P., Campbell, B. M., et al. (2014). Climate-smart agriculture for food security. *Nature Climate Change*, 4(12), 1068-1072.
5. MINAGRI. (2019). *Plan Nacional de Agricultura Familiar 2019-2021*. Lima: Ministerio de Agricultura y Riego del Perú.
6. OECD. (2020). *Rural Policy Reviews: Linking rural development policies with evidence-based approaches*. Paris: OECD Publishing.
7. Ostrom, E. (2009). A general framework for analyzing sustainability of social-ecological systems. *Science*, 325(5939), 419-422.
8. Pretty, J., Benton, T. G., Bharucha, Z. P., et al. (2018). Global assessment of agricultural system redesign for sustainable intensification. *Nature Sustainability*, 1(8), 441-446.
9. Springmann, M., Clark, M., Mason-D'Croz, D., et al. (2018). Options for keeping the food system within environmental limits. *Nature*, 562(7728), 519-525.
10. Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *PNAS*, 108(50), 20260-20264.
11. van Ittersum, M. K., van Bussel, L. G., Wolf, J., et al. (2016). Can sub-Saharan Africa feed itself? *PNAS*, 113(52), 14964-14969.
12. Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big Data in Smart Farming—A review. *Agricultural Systems*, 153, 69-80.
13. World Bank. (2019). *Peru: Managing Agricultural Production Risk*. Washington D.C.: World Bank Group.
14. Zhang, L., Dabipi, I. K., & Brown, W. L. (2021). Internet of Things applications for agriculture: A systematic review. *Computers and Electronics in Agriculture*, 187, 106168.
15. Zimmerer, K. S., de Haan, S., Jones, A. D., et al. (2019). The biodiversity of food and agriculture in the Anthropocene. *Annual Review of Environment and Resources*, 44, 137-168.