

# Universidad Católica Boliviana

## Modelado Dinámica de Sistemas y Simulación



UNIVERSIDAD  
**CATÓLICA**  
BOLIVIANA

### Examen Primera Sesión: Sistema de Análisis y Predicción de Rendimiento de Futbolistas

Este documento contiene las especificaciones técnicas detalladas, rúbricas de evaluación y guías de implementación para el sistema integrado con modelos avanzados.

## Índice

<b>1. Información General del Proyecto</b>	<b>3</b>
1.1. Objetivos Pedagógicos . . . . .	3
1.2. Arquitectura del Sistema . . . . .	3
<b>2. Instrucciones Detalladas por Día</b>	<b>4</b>
2.1. Día 1: Preprocesamiento y Modelos Lineales . . . . .	4
2.1.1. Objetivos Específicos . . . . .	4
2.1.2. Tareas por Estudiante . . . . .	4
2.1.3. Entregables Día 1 . . . . .	6
2.2. Día 2: Redes Neuronales Multicapa . . . . .	7
2.2.1. Arquitecturas a Implementar . . . . .	7
2.2.2. Tareas por Estudiante . . . . .	7
2.2.3. Entregables Día 2 . . . . .	8
2.3. Día 3: Integración Total y Sistema Dinámico . . . . .	9
2.3.1. Modelo Dinámico con Integración Numérica . . . . .	9
2.3.2. Tareas por Estudiante . . . . .	9
2.3.3. Entregables Día 3 . . . . .	11
<b>3. Rúbricas de Evaluación Detalladas</b>	<b>12</b>
3.1. Rúbrica General (100 puntos) . . . . .	12
3.2. Rúbricas Específicas por Modelo . . . . .	13
3.2.1. Regresión Lineal (30 puntos) . . . . .	13
3.2.2. Redes Neuronales (30 puntos) . . . . .	13
3.2.3. Integración Numérica (20 puntos) . . . . .	13
3.3. Evaluación Individual por Competencias . . . . .	14
3.4. Escala de Calificación Final . . . . .	14

<b>4. Recursos y Material de Apoyo</b>	<b>16</b>
4.1. Referencias Teóricas por Tema . . . . .	16
<b>5. Consideraciones Finales y Evaluación</b>	<b>16</b>
5.1. Criterios de Éxito del Proyecto . . . . .	16
5.2. Distribución del Puntaje Final . . . . .	17
5.3. Lineamientos para la Presentación Final . . . . .	17

## Información General del Proyecto

### Objetivos Pedagógicos

1. Integrar múltiples técnicas de modelado en un sistema coherente
2. Aplicar integración numérica a problemas de evolución temporal
3. Implementar filtros de Kalman para estimación de estado
4. Diseñar y entrenar modelos de machine learning
5. Desarrollar habilidades de trabajo en equipo y gestión de proyectos

### Arquitectura del Sistema

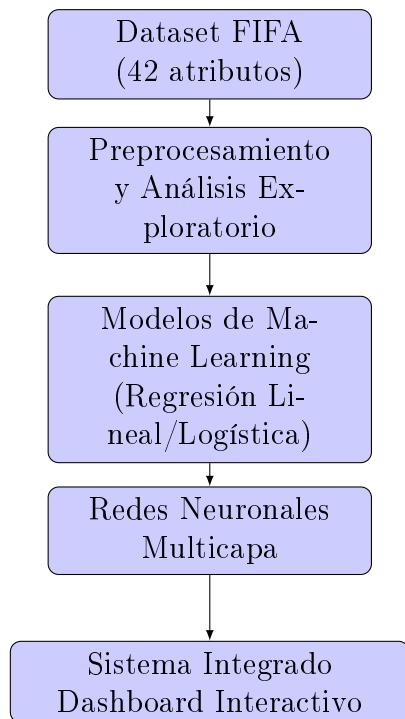


Figura 1: Arquitectura completa del sistema integrado

## Instrucciones Detalladas por Día

### Día 1: Preprocesamiento y Modelos Lineales

#### Objetivos Específicos

- Limpieza y preparación del dataset FIFA
- Implementación de regresión lineal múltiple
- Implementación de regresión logística multiclasificación
- Análisis de importancia de variables

#### Tareas por Estudiante

##### Estudiante A: Ingeniero de Datos

###### 1. Limpieza de datos:

- Manejo de valores nulos usando medianas por posición
- Codificación one-hot para variables categóricas
- Normalización para atributos numéricos (0-100)
- Detección y tratamiento de outliers (atípicos)

###### 2. Creación de variables derivadas:

- Score físico: promedio de acceleration, sprint\_speed, stamina, strength
- Score técnico: promedio de ball\_control, dribbling, short\_passing
- Score mental: promedio de positioning, vision, reactions
- Edad estimada basada en fecha de registro

##### Estudiante B: Especialista en Regresión Lineal

###### 1. Modelo de regresión lineal múltiple:

- Variable objetivo: overall\_rating
- Variables predictoras: 15 atributos clave seleccionados
- Implementar con validación cruzada k-fold ( $k=5$ )
- Calcular RMSE, MAE,  $R^2$

###### 2. Análisis de coeficientes:

- Identificar características más influyentes

- Calcular importancia relativa de cada atributo
- Analizar multicolinealidad
- Volver a entrenar hasta quedar con las mejores métricas.

### Estudiante C: Especialista en Regresión Logística

#### 1. Clasificación de posiciones:

- Crear variable objetivo: posición (portero, defensa, medio, atacante)
- Basado en atributos específicos por posición
- Balancear dataset si es necesario

#### 2. Modelo One-vs-Rest:

- Implementar regresión logística multiclas
- Usar regularización para evitar overfitting
- Calcular matriz de confusión y métricas por clase.

#### 3. Análisis de probabilidades:

- Calcular probabilidades por posición
- Identificar jugadores "híbridos" (múltiples posiciones posibles)

### Estudiante D: Coordinador y Validador

#### 1. Integración de modelos:

- Crear pipeline unificado de preprocesamiento
- Integrar modelos de regresión lineal y logística
- Validar consistencia entre modelos

#### 2. Validación estadística:

- Test de normalidad de residuos
- Test de homocedasticidad
- Validación de supuestos de modelos lineales

#### 3. Documentación y presentación:

- Documentar hallazgos clave
- Preparar presentación de resultados
- Crear dashboard básico con visualizaciones

### Entregables Día 1

- Dataset limpio en formato SQLite
- Modelos de regresión lineal entrenados y evaluados
- Modelo de regresión logística entrenado
- Reporte de importancia de variables
- Presentación de 15 minutos

## Día 2: Redes Neuronales Multicapa

### Arquitecturas a Implementar

#### Red 1: Predicción de Potencial Máximo

- **Entrada:** 20 características (atributos )
- **Arquitectura:** 20-256-128-64-1 neuronas
- **Funciones activación:** ReLU en capas ocultas, lineal en salida
- **Regularización:**  $\lambda$
- **Salida:** Potencial máximo alcanzable (potential)

#### Red 2: Clasificación de Perfil de Jugador

- **Entrada:** 15 atributos clave
- **Arquitectura:** 15-256-128-7 neuronas
- **Funciones activación:** ReLU ocultas, softmax salida
- **Salida:** Probabilidades para 7 posiciones

### Tareas por Estudiante

#### Estudiante D: Diseño de Arquitecturas

##### 1. Selección de hiperparámetros:

- Número de capas y neuronas por capa
- Funciones de activación óptimas
- Tasa de aprendizaje y optimizador

##### 2. Prevención de overfitting:

- Regularización  $\lambda$

##### 3. Normalización de datos:

- Escalado apropiado para redes neuronales
- Manejo de variables categóricas

#### Estudiante A: Entrenamiento y Optimización

**1. División de datos:**

- 70 % entrenamiento, 15 % validación, 15 % test
- Balanceo por posición/clase
- Estratificación para mantener distribuciones

**2. Proceso de entrenamiento:**

- Implementación con modelo codificado
- Verificar mejores hiperparámetros

**Estudiante B: Evaluación e Interpretabilidad****1. Métricas de evaluación:**

- Regresión: MAE, RMSE, R<sup>2</sup>, Error máximo
- Clasificación: Accuracy, Precision, Recall, F1, AUC-ROC

**2. Análisis de errores:**

- Matrices de confusión detalladas
- Análisis de residuos por tipo de jugador
- Identificación de patrones de error sistemático

**3. Interpretabilidad:**

- SHAP values para importancia de características
- Visualización de activaciones de capas

**Estudiante C: Integración y Validación****1. Validación cruzada:**

- K-fold cross validation (k=5)
- Stratified k-fold para clasificación

**Entregables Día 2**

- 3 redes neuronales entrenadas y guardadas
- Reporte completo de métricas de evaluación
- Análisis de importancia de características (SHAP)
- Comparativa con modelos baseline
- Presentación de 20 minutos

## Día 3: Integración Total y Sistema Dinámico

### Modelo Dinámico con Integración Numérica

**Sistema de Ecuaciones Diferenciales** Modelamos la evolución de un futbolista como:

$$\begin{aligned}\frac{dF}{dt} &= \alpha_F E_F e^{-\frac{(A-A_{opt})^2}{2\sigma^2}} - \beta_F \max(0, A - 30)F \\ \frac{dT}{dt} &= \alpha_T E_T + \gamma_{FT} F - \beta_T \max(0, A - 32)T \\ \frac{dM}{dt} &= \alpha_M E_M + \delta(F + T) - \beta_M \max(0, A - 35)M \\ \frac{dR}{dt} &= w_F \frac{dF}{dt} + w_T \frac{dT}{dt} + w_M \frac{dM}{dt} + \epsilon \\ \frac{dA}{dt} &= \frac{1}{365}\end{aligned}$$

Donde:

- $F, T, M$ : Puntuaciones física, técnica, mental (0-1)
- $R$ : Overall rating (40-100)
- $A$ : Edad (años)
- $E_F, E_T, E_M$ : Intensidad de entrenamiento (0-1)
- $\alpha$ : Tasas de aprendizaje,  $\beta$ : Tasas de decaimiento
- $\gamma, \delta$ : Factores de sinergia entre atributos

### Métodos de Integración Numérica

- **RK4**: Alta precisión, estable

### Tareas por Estudiante

#### Estudiante C: Modelado de Ecuaciones Diferenciales

##### 1. Definición de parámetros:

- Calibrar  $\alpha, \beta, \gamma$  con datos históricos (modelos ML)
- Estimar  $A_{opt}$  (edad óptima) por posición
- Definir pesos  $w_F, w_T, w_M$  para rating

**2. Modelado de entrenamiento:**

- Definir regímenes de entrenamiento típicos
- Modelar efecto de lesiones (reducción temporal de capacidades)
- Incluir efecto de competiciones (fatiga acumulada)

**Estudiante B: Implementación Numérica****1. Implementación de solvers:**

- Implementar RK4 manualmente para comprensión

**Estudiante A: Simulación de Escenarios****1. Escenarios de desarrollo:**

- "Joven promesa": Alto potencial, desarrollo rápido
- "Lento pero seguro": Desarrollo constante
- "Early bloomer": Pico temprano, declive rápido
- "Late bloomer": Desarrollo tardío

**2. Análisis de sensibilidad:**

- Variar intensidad de entrenamiento
- Simular efecto de lesiones graves
- Estudiar impacto de diferentes regímenes

**3. Optimización de trayectorias:**

- Encontrar régimen óptimo para maximizar potencial
- Balancear desarrollo físico vs técnico
- Considerar riesgo de lesiones vs beneficios

**Estudiante D: Sistema Integrado Final****1. Integración de todos los componentes:**

- Usar predicciones de redes como input para simulaciones
- Crear pipeline completo: datos → redes → simulación

**2. Dashboard interactivo:**

- Visualizaciones interactivas de simulaciones
- Interface para modificar parámetros en tiempo real

**3. Casos de estudio completos:**

- Analizar 5 jugadores reales con sistema completo
- Generar reportes personalizados con recomendaciones
- Presentar resultados integrados

### Entregables Día 3

- Sistema integrado completamente funcional
- Dashboard interactivo web-based
- Reporte técnico completo (15-20 páginas)
- 5 casos de estudio documentados
- Presentación final de 25 minutos

## Rúbricas de Evaluación Detalladas

### Rúbrica General (100 puntos)

Criterio	Puntos	Excelente (90-100 %)	Suficiente (60-70 %)
<b>1. Calidad Técnica (40 pts)</b>	40		
Implementación correcta	10	Todos los componentes funcionan perfectamente	Componentes básicos funcionan
Fundamentación teórica	10	Explicación matemática completa	Explicación básica presente
Validación rigurosa	10	Múltiples métodos de validación	Validación básica realizada
Manejo de errores	5	Robustez frente a todos los casos	Manejo básico de errores
Optimización	5	Código altamente optimizado	Código funcional pero no optimizado
<b>2. Trabajo en Equipo (20 pts)</b>	20		
Colaboración efectiva	5	Comunicación excelente, apoyo mutuo	Comunicación básica
Distribución equilibrada	5	Roles claros, responsabilidades compartidas	Algunos desbalances
Integración de componentes	5	Integración perfecta, coherencia total	Integración básica lograda
Resolución de conflictos	5	Manejo constructivo, soluciones creativas	Conflictos resueltos básicamente
<b>3. Documentación (20 pts)</b>	20		
Compleitud	5	Todas las secciones completas y detalladas	Secciones principales completas
Claridad	5	Explicaciones excepcionalmente claras	Explicaciones comprensibles
Organización	5	Estructura lógica y profesional	Estructura básicamente organizada
Reproducibilidad	5	Instrucciones exactas para replicación	Instrucciones generales presentes
<b>4. Presentación (20 pts)</b>	20		
Estructura presentación	5	Flujo lógico, atractivo visual	Estructura básica presente
Calidad visual	5	Gráficos profesionales, diseño excelente	Gráficos básicos presentes
Comunicación oral	5	Claridad, confianza, dominio total	Comunicación básicamente clara
Respuestas preguntas	5	Respuestas profundas, conocimiento completo	Respuestas adecuadas
<b>TOTAL</b>	<b>100</b>		

Cuadro 1: Rúbrica general de evaluación

## Rúbricas Específicas por Modelo

### Regresión Lineal (30 puntos)

Criterio	Puntos	Excelente (5)	Suficiente (3)
Selección de variables	4	Selección óptima con justificación	Selección básica adecuada
Validación de supuestos	4	Todos los supuestos verificados	Supuestos principales verificados
Métricas de performance	4	$RMSE < 2.0$ , $R^2 > 0.85$	$RMSE < 3.0$ , $R^2 > 0.70$
Análisis de residuos	4	Análisis completo, patrones identificados	Análisis básico realizado
Interpretación coeficientes	4	Interpretación profunda, insights valiosos	Interpretación básica correcta
<b>Subtotal</b>	<b>20</b>		

### Redes Neuronales (30 puntos)

Criterio	Puntos	Excelente (5)	Suficiente (3)
Diseño arquitectura	4	Arquitectura óptima, bien justificada	Arquitectura básica funcional
Prevención overfitting	4	Múltiples técnicas, efectivas	Técnicas básicas implementadas
Métricas de evaluación	4	Métricas completas, comparativas	Métricas principales calculadas
Interpretabilidad	4	SHAP, análisis profundo	Interpretación básica realizada
Validación cruzada	4	CV anidada, resultados robustos	K-fold básico implementado
<b>Subtotal</b>	<b>20</b>		

### Integración Numérica (20 puntos)

Criterio	Puntos	Excelente (5)	Suficiente (3)
Modelado ecuaciones	4	Ecuaciones realistas, bien parametrizadas	Ecuaciones básicas correctas
Implementación solver	4	Método adaptativo, alta precisión	Solver básico funcionando
Ánalysis estabilidad	4	Ánalysis completo de estabilidad	Estabilidad básica verificada
Simulación escenarios	4	Múltiples escenarios relevantes	Escenarios básicos simulados
Validación resultados	4	Validación rigurosa con datos reales	Validación básica realizada
<b>Subtotal</b>	<b>20</b>		

## Evaluación Individual por Competencias

Competencia	Puntos	Descripción
<b>Competencia Técnica Individual</b>	45	
Calidad del código en su área	15	Limpio, eficiente, bien documentado
Documentación técnica de su trabajo	10	Completa, clara, profesional
Innovación y creatividad	10	Soluciones originales, mejoras creativas
Resolución de problemas técnicos	10	Capacidad para superar obstáculos
<b>Contribución al Trabajo en Equipo</b>	35	
Cumplimiento responsabilidades	10	Tareas completadas a tiempo y calidad
Colaboración y apoyo	10	Ayuda activa a compañeros, trabajo conjunto
Comunicación efectiva	10	Claridad, frecuencia, calidad comunicación
Flexibilidad y adaptación	5	Adaptación a cambios, roles flexibles
<b>Calidad de Entregables</b>	20	
Completitud entregables	10	Todos los entregables completos y correctos
Calidad presentación	5	Presentación profesional, atractiva
Puntualidad	5	Entregas a tiempo según cronograma
<b>TOTAL INDIVIDUAL</b>	<b>100</b>	

Cuadro 2: Rúbrica de evaluación individual

## Escala de Calificación Final

$$\text{Calificación Final} = 0,6 \times \text{Promedio Individual} + 0,4 \times \text{Calificación Proyecto}$$

- **Excelente (90-100):** Sistema completo funcional, integración impecable, documentación excepcional
- **Muy Bueno (80-89):** Sistema funcional con pequeñas mejoras posibles, buena integración
- **Bueno (70-79):** Sistema básicamente funcional, algunos componentes mejorables

- **Suficiente (60-69):** Sistema funciona parcialmente, necesita mejoras significativas
- **Insuficiente (<60):** Sistema no funciona o falta integración esencial

## Recursos y Material de Apoyo

### Referencias Teóricas por Tema

#### Regresión Lineal y Logística:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*
- Agresti, A. (2012). *Categorical Data Analysis*

#### Redes Neuronales:

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*
- Chollet, F. (2018). *Deep Learning with Python*

#### Integración Numérica:

- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical Recipes*
- Hairer, E., Nørsett, S. P., & Wanner, G. (1993). *Solving Ordinary Differential Equations I*

#### Análisis Deportivo:

- Pappalardo, L., et al. (2019). *A public data set of spatio-temporal match events in soccer competitions*
- Duch, J., Waitzman, J. S., & Amaral, L. A. N. (2010). *Quantifying the performance of individual players*

## Consideraciones Finales y Evaluación

### Criterios de Éxito del Proyecto

El proyecto se considerará exitoso si cumple:

1. **Funcionalidad completa:** Sistema procesa datos y genera predicciones coherentes
2. **Integración efectiva:** Componentes trabajan conjuntamente de forma fluida
3. **Validación empírica:** Predicciones con error aceptable ( $\text{RMSE} < 3.0$  para rating)
4. **Escalabilidad:** Sistema puede procesar múltiples jugadores eficientemente
5. **Usabilidad:** Interfaz clara y documentación suficiente para uso por terceros
6. **Innovación:** Aportes creativos en modelado o implementación

## Distribución del Puntaje Final

Componente	Ponderación	Puntos
Puntaje individual (promedio)	60 %	60
Calidad del sistema integrado	20 %	20
Documentación y presentación	15 %	15
Innovación y creatividad	5 %	5
<b>TOTAL</b>	<b>100 %</b>	<b>100</b>

## Lineamientos para la Presentación Final

**Duración:** 20 minutos (15 presentación + 5 preguntas)

**Estructura sugerida:**

1. **Introducción (3 min):** Problema, objetivos, motivación
2. **Dataset y preprocesamiento (4 min):** Desafíos, decisiones técnicas
3. **Modelos lineales (4 min):** Regresión lineal y logística, resultados
4. **Redes neuronales (4 min):** Arquitecturas, entrenamiento, evaluación
5. **Sistema integrado (3 min):** Demostración, casos de estudio
6. **Conclusiones (2 min):** Lecciones aprendidas, trabajo futuro

**Requisitos de participación:**

- Todos los miembros deben hablar durante la presentación
- Incluir demostración en vivo del sistema
- Preparar respuestas para preguntas técnicas
- Usar visualizaciones descriptivas y claras

- Todos saben el completo del trabajo

**¡Éxito en la implementación del sistema integrado!**