

## Índice de contenidos.

<b>Resumen.....</b>	<b>5</b>
<b>Introducción.....</b>	<b>6</b>
1.1 Sistema inmunitario.....	6
1.2 Sistema inmune innato y adaptativo.....	6
2. Reconocimiento de epítomos para linfocitos T y B.....	10
3. Descriptores moleculares.....	10
4. Machine learning.....	11
4.1. Modelos de predicción mediante inteligencia artificial.....	12
<b>Hipótesis y objetivos.....</b>	<b>13</b>
1. Problema.....	13
2. Hipótesis.....	13
3. Objetivos.....	13
3.1. Objetivo General.....	13
3.2. Objetivos específicos.....	13
<b>Materiales y metodología.....</b>	<b>14</b>
1.1 Identificar y analizar bases de datos para epítomos proteicos.....	14
1.2 Implementar descriptores moleculares para caracterizar proteína a nivel de estructura primaria y terciaria.....	16
1.3 Limpieza, estandarización y selección de atributos óptimos para el set de datos.....	19
1.4 Implementar y comparar modelos predictivos en inteligencia artificial, evaluando exactitud y sensibilidad.....	22
<b>Resultados.....</b>	<b>25</b>
Obj 1. Identificar y analizar bases de datos para epítomos proteicos.....	25
Obj 2. Implementar descriptores moleculares para caracterizar proteína a nivel de estructura primaria y terciaria.....	25
Obj 3. Limpieza, estandarización y selección de atributos óptimos para el set de datos.....	26
Obj. 4 Implementar y comparar modelos predictivos en inteligencia artificial, evaluando exactitud y sensibilidad.....	27
<b>Discusión.....</b>	<b>37</b>
<b>Referencias.....</b>	<b>43</b>

## Índice de Figuras.

### Introducción.

- 1. Sistema inmunitario innato y adaptativo..... 8
- 2. Eliminación de patógenos por linfocito T..... 9
- 3. Modelos de aprendizaje..... 11

### Materiales y metodología.

- 4. Filtros para la búsqueda de epítomos..... 14
- 5. Alineamiento múltiple de epítomos..... 16
- 6. Fragmentación de secuencia..... 17
- 7. Lectura del conjunto de datos..... 18
- 8. Selección de atributos..... 20
- 9. Validación cruzada..... 21
- 10. Método Random forest..... 22
- 11. Matriz de confusión..... 23

### Resultados.

- 12. Archivo de salida script 1..... 25
- 13. Normalización de datos..... 27
- 14. Balanceo de clases..... 29
- 15. Selección de árboles de decisión..... 30
- 16. Análisis de selección de atributos..... 31
- 17. Dispersión de atributos..... 32

## Índice de tablas.

### Resultados.

- 1. Resultados del primer ciclo.....	28
- 2. Resultados del segundo ciclo.....	29
- 3. Predicción de casos aislados.....	33
- 4. Probabilidad de epítomos reconocidos.....	34
- 5. Predicción de epítomos en 6BKN por RF.....	34
- 6. Predicción de epítomos en 7JM3 por RF.....	35

### Discusión.

- 7. Comparación de modelos.....	37
- 8. Predicción de epítomos en 6BKN por DTU.....	38
- 9. Predicción de epítomos en 7JM3por DTU.....	39