

**Universidad Tecnológica**

**Centroamericana**

Facultad de Ingeniería

**Mini Proyecto 2.**

**Docente:**

PhD. Kenny Mauricio Dávila.

**Asignatura:**

Sistemas Inteligentes.

**Presentado por:**

Ingrid Domínguez 11711355

Claudia Cortés 11711357

**Tegucigalpa, M.D.C,**

***Diciembre de 2020***

Contenido

[1. Introducción. 3](#_Toc58282156)

[2. Descripción de la implementación. 3](#_Toc58282157)

[3. Resultados y Estadísticas: Parte 1. 4](#_Toc58282158)

[4. Resultados y Estadísticas: Parte 2. 8](#_Toc58282159)

[5. Resultados y Estadísticas: Parte 3. 8](#_Toc58282160)

[6. Resultados y Estadísticas: Parte 4. 8](#_Toc58282161)

[7. Análisis Resumido 8](#_Toc58282162)

[8. Dificultades encontradas. 8](#_Toc58282163)

[9. Conclusiones. 8](#_Toc58282164)

# Introducción.

# Descripción de la implementación.

* Comandos y recomendaciones para la ejecución:
  1. Puesto que el programa requiere de librerías especificas que necesitan instalarse, se preparó un script llamado “***septup.py***” el cual encontrará en la carpeta principal del proyecto. Ejecute el archivo “***setup.py***” sin ningún parámetro en la línea de comando para instalar las librerías necesarias.
  2. La ejecución del archivo “***mp\_2\_parte\_1.py***” requiere la existencia de una carpeta llamada ***“GRAFICAS”,*** dicha carpeta ya se encuentra creada en la carpeta principal del programa, por lo cual se recomienda no borrar dicha carpeta.
  3. Las estadísticas y graficas generadas en la parte 1 (mp\_2\_parte\_1.py) se generarán automáticamente en la carpeta “GRAFICAS”.
  4. Los parámetros necesarios para la ejecución de cada archivo son los mismos que los especificados en la descripción del proyecto.
* librerías especificas utilizadas:
  + Parte1:

1. *Pandas*: Para el manejo de archivos en general.
2. *xlsxwriter,xlrd,openpyxl*: Perteneciente a la librería Pandas, se utiliza para la gestión de archivos en Excel, este fue utilizado para reportar los resultados de las estadísticas de la parte 1 en un documento con extensión .xls.
3. *Numpy*: Para dar soporte a la creación de vectores/matrices.
4. *matplotlib.pyplot*: Para la generación de graficas.

* Parte 2:
  1. RandomForestClassifier
  2. make\_classification
* Parte3:
* Parte4:

# Resultados y Estadísticas: Parte 1.

*Clasificación de los atributos.*

|  |  |  |
| --- | --- | --- |
| Tipo de Atributo | Cantidad | Descripción |
| Continuo | 4 | Por su naturaleza se clasificaron como discretos los atributos:   * Plaquetas * Linfocitos * Hematocritos * Leucocitos |
| Discreto | 17 | Se consideran discretos los demás atributos no listados en el inciso anterior. |

1. *Estadísticas para valores discretos.*

*Nota: Los valores a continuación presentados fueron obtenidos del dataset completo\_train\_synth\_dengue.csv”*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **sexo** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| M | 2914 | 1028 | 510 | 502 |
| F | 3068 | 952 | 532 | 494 |

Tabla . Resultados estadísticos Sexo

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **nauseas** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| Si | 1755 | 427 | 218 | 610 |
| No | 4227 | 1553 | 824 | 386 |

Tabla . Resultados estadísticos Nauseas

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **vomitos** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| No | 3670 | 1697 | 625 | 241 |
| Si | 1835 | 103 | 35 | 10 |
| Persistente | 477 | 180 | 382 | 745 |

Tabla . Resultados estadísticos vomitos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **rash** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| Si | 323 | 501 | 265 | 465 |
| No | 5659 | 1479 | 777 | 531 |

Tabla . Resultados estadísticos rash

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **artralgias** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| Si | 2441 | 1583 | 851 | 792 |
| No | 3541 | 397 | 191 | 204 |

Tabla . Resultados estadísticos artralgias

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **mialgias** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| Si | 3604 | 1585 | 831 | 818 |
| No | 2378 | 395 | 211 | 178 |

Tabla . Resultados estadísticos mialgias

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **dias\_fiebre** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| 4 | 858 | 389 | 187 | 361 |
| 3 | 1095 | 1211 | 674 | 324 |
| 1 | 2419 | 0 | 0 | 0 |
| 5 | 0 | 380 | 181 | 311 |
| 2 | 1610 | 0 | 0 | 0 |

Tabla . Resultados estadísticos días con fiebre

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **dias\_ultima\_fiebre** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| 1 | 1144 | 305 | 158 | 222 |
| 0 | 1753 | 1116 | 575 | 252 |
| 2 | 931 | 285 | 142 | 258 |
| 4 | 0 | 82 | 57 | 83 |
| 3 | 2154 | 192 | 110 | 181 |

Tabla . Resultados estadísticos días desde la última fiebre

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **prueba\_torniquete** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| Positiva | 194 | 421 | 222 | 741 |
| NA | 5067 | 497 | 262 | 244 |
| Negativa | 721 | 1062 | 558 | 11 |

Tabla . Resultados estadísticos prueba torniquete

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **dolor\_abdominal** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| No | 3638 | 1799 | 671 | 273 |
| Si | 2344 | 181 | 371 | 723 |

Tabla . Resultados estadísticos dolor abdominal

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **acumulacion\_fluidos** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| No | 5680 | 1785 | 658 | 251 |
| Si | 302 | 195 | 384 | 745 |

Tabla . Resultados estadísticos acumulación fluidos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **sangrado\_mucosas** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| No | 5918 | 1867 | 887 | 197 |
| Si | 64 | 113 | 155 | 799 |

Tabla . Resultados estadísticos sangrado en la mucosa

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **hemorragia** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| No | 4987 | 1980 | 1042 | 197 |
| Si | 995 | 0 | 0 | 799 |

Tabla .Resultados estadísticos hemorragia

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **shock** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| No | 5521 | 1980 | 1042 | 592 |
| Si | 461 | 0 | 0 | 404 |

Tabla .Resultados estadísticos shock

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **letargia** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| No | 5658 | 1797 | 652 | 307 |
| Si | 324 | 183 | 390 | 689 |

Tabla .Resultados estadísticos letargia

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **irritabilidad** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| No | 5682 | 1969 | 1011 | 923 |
| Si | 300 | 11 | 31 | 73 |

Tabla .Resultados estadísticos irritabilidad

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **hepatomegalia** | **No Dengue** | **No signos Alerta** | **Signos Alerta** | **Dengue Grave** |
| No | 5796 | 1832 | 652 | 265 |
| Si | 186 | 148 | 390 | 731 |

Tabla .Resultados estadísticos hepatomegalia

1. *Estadísticas para valores continuos:*

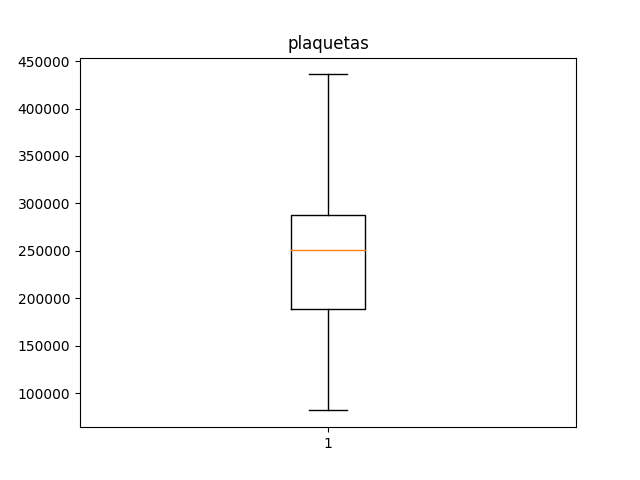


Ilustración Grafica de caja para prueba de laboratorio de plaquetas

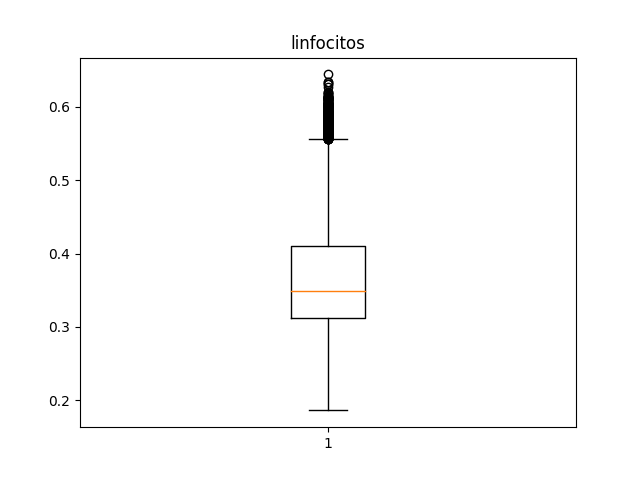


Ilustración Grafica de caja para prueba de laboratorio de linfocitos

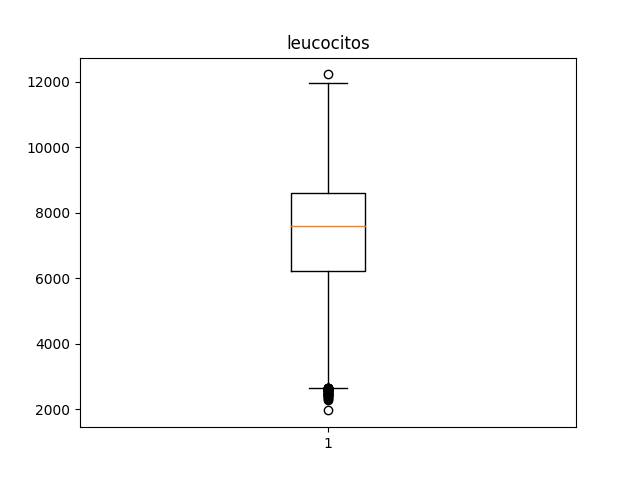


Ilustración . Grafica de caja para prueba de laboratorio de Leucocitos

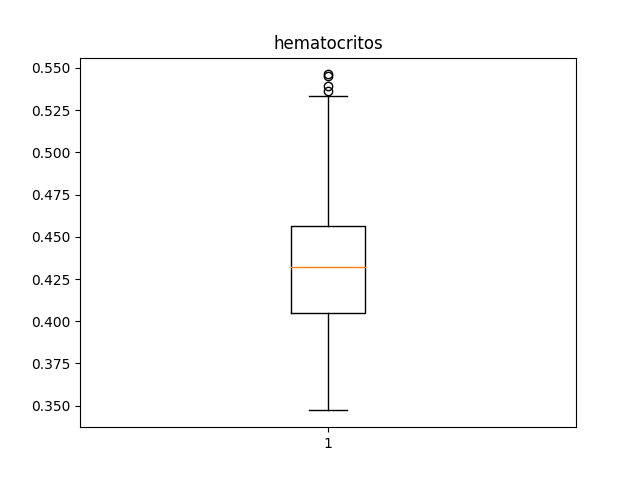


Ilustración Grafica de caja para prueba de laboratorio de hematocritos

1. *Análisis de resultados:*

Luego del análisis de los datos obtenidos de la generación de tablas y graficas de los atributos discretos y continuos, se observo un patrón de comportamiento en los datos de prueba, donde si bien la clase mayoritaria era “No dengue”, las clases minoritarias “dengue no grave y sin signos de alarma”, “dengue no grave con signos de alarma” y “dengue grave” tienen gran importancia al momento del diagnóstico médico.

Algunos atributos mostraban tener importancia dentro del diagnóstico al tomar determinados valores, esto se afirma puesto que en ellos se podía observar una considerable diferencia entre la cantidad de personas que no tenían dengue contra los que si tenían algún tipo de dengue independientemente de la gravedad de este. Algunos ejemplos de estos casos y por lo cual se consideran como atributos de mayor importancia son:

|  |  |  |
| --- | --- | --- |
| Atributo | Valor | Porcentaje de diferencia |
| Prueba Torniquete | Positiva | 87.71% de las personas que dieron “positivo” a la prueba del torniquete tenían algún tipo de dengue. |
| Acumulación de fluidos | Si | 81.43% de las personas que si habían tenido retención de líquidos tenían algún tipo de dengue. |
| Días con fiebre | 3 | 66.86% de las personas que respondieron habían tenido 3 días de fiebre tenían algún tipo de dengue. |
| Hepatomegalia | Si | 87.21% de las personas que si habían sufrido alteraciones en el tamaño del hígado tenían algún tipo de dengue. |
| Sangrado en las mucosas | Si | 94.34% de las personas que tenían sangrado en la mucosidad tenían algún tipo de dengue.  *Nota: La mayoría de las personas no presentaba sangrado en la mucosidad* |

Estos resultados son en base a los atributos discretos, de los atributos continuos se deduce que los más importantes son: XXXXX porque xxxxx y en general y en base a el análisis del grupo se menciona que los atributos mas importantes son:xxxxxxx, xxxxxxx, xxxxx, xxxx

Por otro lado, los atributos que se consideran menos relevantes ya que no muestran información concluyente son: irritabilidad, schock, hemorragia y el sexo.

# Resultados y Estadísticas: Parte 2.

# Resultados y Estadísticas: Parte 3.

# Resultados y Estadísticas: Parte 4.

# Análisis Resumido

1. . ¿Qué clasificador le dio los mejores resultados en validación?

Después de realizar un análisis individual y comparara los resultados de los diferentes clasificadores implementados en este proyecto se llega a la conclusión de que el XXX dio los mejores resultados en la validación de los datos utilizados.

1. ¿Qué clasificador le dio los mejores resultados en los datos de prueba?

Para los datos de prueba que se proporcionaron y se utilizaron en los diferentes clasificadores, se considera que el mejor resultado lo tuvo el XXX ya que en comparación a los otros clasificadores sus métricas mostraban una mayor (precisión? (x vs x).

1. ¿Qué clasificador prefiere y por qué razones?

Luego de implementar y poner en comparativa los diferentes clasificadores, el grupo llega a la conclusión de que se prefiere el XXX esto es por motivos de rendimiento, facilidad de compresión de predicciones y entendimiento del funcionamiento del mismo.

# Dificultades encontradas.

# Conclusiones.