

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

“MAGÍSTER EN ESTADÍSTICA APLICADA”

TEMA:

HERRAMIENTA DE VISUALIZACIÓN INTERACTIVA, PARA UN PROCESO DE DATA MINING BASADO EN BIPLLOT, CON LENGUAJE DE PROGRAMACIÓN WEB: APLICACIÓN AL ANÁLISIS DE PRUEBAS DE LABORATORIO CLÍNICO, EN PACIENTES CON DIAGNÓSTICO DE CÁNCER.

AUTOR:

ING. CRISTIAN ARGELIS TOMALÁ MAZZINI, MSIG

Guayaquil – Ecuador

2024

RESUMEN

El presente trabajo de titulación tiene como finalidad promover el desarrollo de aplicaciones web que permitan analizar las variables y observaciones de diversos conjuntos de datos por medio de la ilustración de gráficos interactivos de la familia Biplot. El análisis del conjunto de datos principal se basa sobre pruebas de laboratorio clínico de pacientes con patología de cáncer de un Hospital Oncológico radicado en la ciudad de Guayaquil - Ecuador.

En el capítulo I, se abordará el entorno sanitario en las instituciones prestadoras de servicios de salud en el Ecuador, se analizará la problemática de comprender de forma visual la relación de la historia clínica del paciente almacenados en diversas bases de datos teniendo en cuenta el volumen de información considerable, y se elaborará una propuesta para solucionar el problema de observar la relación de la historia clínica del paciente por medio de una aplicación web interactiva.

En el capítulo II, se define el marco conceptual de los métodos de la familia Biplot, software para graficar los métodos de la familia Biplot, librerías de R para graficar los métodos de la familia Biplot, sistemas de información hospitalarios, sistemas de información de laboratorio clínico, además de lenguajes de programación de código abierto para el desarrollo de aplicaciones web, y herramientas de código abierto para aplicaciones web, que permitan diseñar y desarrollar gráficos de la familia Biplot.

En el capítulo III, se ilustra la metodología que se abordará para analizar, diseñar y desarrollar la aplicación web interactiva, haciendo uso de lenguajes y herramientas de programación de código abierto para aplicaciones web. Así mismo, esta metodología permitirá analizar y preparar el conjunto de datos de pacientes con patología de cáncer que se usará en el entorno web interactivo.

En el capítulo IV, se exponen los resultados del análisis sobre el conjunto de datos de la relación de la historia clínica de pacientes con patología de cáncer haciendo uso del entorno web interactivo en el que se usarán las funcionalidades de la aplicación web y se presentarán los gráficos de los métodos de la familia Biplot.

Finalmente, en el capítulo V se exhiben las conclusiones y recomendaciones del presente trabajo tanto del entorno web interactivo como del conjunto de datos sobre la relación de la historia clínica de pacientes con patología de cáncer.

ABSTRACT

The purpose of this degree work is to promote the development of web applications that allow analyzing the variables and observations of various data sets through the illustration of interactive graphics of the Biplot family. The analysis of the main data set is based on clinical laboratory tests of patients with cancer pathology from an Oncology Hospital located in the city of Guayaquil - Ecuador.

In chapter I, the health environment in the institutions providing health services in Ecuador will be addressed, the problem of visually understanding the relationship of the patient's medical history stored in various databases will be analyzed, taking into account the volume of considerable information, and a proposal will be developed to solve the problem of observing the relationship of the patient's medical history through an interactive web application.

In chapter II, the conceptual framework of the Biplot family methods is defined, software to graph the Biplot family methods, R libraries to graph the Biplot family methods, hospital information systems, laboratory information systems clinical, in addition to open source programming languages for the development of web applications, and open source tools for web applications, which allow the design and development of Biplot family graphics.

Chapter III illustrates the methodology that will be addressed to analyze, design and develop the interactive web application, using open source programming languages and tools for web applications. Likewise, this methodology will allow the analysis and preparation of the data set of patients with cancer pathology that will be used in the interactive web environment.

In chapter IV, the results of the analysis on the data set of the relationship of the clinical history of patients with cancer pathology are presented using the interactive web

environment in which the functionalities of the web application will be used and the results will be presented. graphics of the Biplot family methods.

Finally, Chapter V presents the conclusions and recommendations of this work, both from the interactive web environment and from the data set on the relationship between the clinical history of patients with cancer pathology.

DEDICATORIA

Dedico este proyecto de titulación a mi esposa Daryna, por apoyarme incondicionalmente en todo momento.

A mis hijos Mateo, Josué y Adriana, para que en el futuro les sirva como ejemplo de superación y trabajo constante.

A mi padre Isidro y mi madre Hilda, quienes me enseñaron con su ejemplo a trabajar duramente, y son mi cimiento en la construcción de mi carrera profesional.

AGRADECIMIENTO

Principalmente a Dios, por concederme la sabiduría, discernimiento y salud para la culminación de este proyecto de titulación y sobre todo por darme la oportunidad de formar nuevas amistades. A los docentes quienes me impartieron su conocimiento y parte de su experiencia profesional.

Además, a esta célebre universidad que me abrió sus puertas para cumplir un objetivo más de mi carrera profesional.

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación, me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo, corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Cristian Argelis Tomalá Mazzini

TRIBUNAL DE GRADUACIÓN

PhD. Mariela González Narváez
PRESIDENTE

PhD. Johny Pambabay Calero
TUTOR

Mgtr. Francisco Moreira Villegas
DOCENTE EVALUADOR

TABLA DE CONTENIDO

RESUMEN	II
ABSTRACT	IV
DEDICATORIA	VI
AGRADECIMIENTO	VII
DECLARACIÓN EXPRESA.....	VIII
TRIBUNAL DE GRADUACIÓN.....	IX
TABLA DE CONTENIDO.....	X
ABREVIATURAS O SIGLAS.....	XV
LISTADO DE FIGURAS.....	XVII
LISTADO DE TABLAS	XIX
CAPÍTULO 1	1
1. INTRODUCCIÓN	1
1.1 ANTECEDENTES	1
1.2 DESCRIPCIÓN DEL PROBLEMA.....	2
1.3 SOLUCIÓN PROPUESTA	4
1.4 OBJETIVO GENERAL.....	4
1.5 OBJETIVOS ESPECÍFICOS	5
1.6 ALCANCE.....	5
CAPÍTULO 2	7
2. MARCO TEÓRICO	7
2.1 BIPLOTS CLASICOS.....	7
2.1.1 JK-BIPLLOT	9

2.1.2	GH-BILOT	10
2.1.3	SQRT-BILOT	10
2.2	HJ-BILOT	10
2.3	CDBILOT	11
2.4	SPARSE BILOT	12
2.5	SOFTWARE BILOT	14
2.5.1	IBM SPSS STATISTICS	15
2.5.2	STATISTICAL ANALYSIS SOFTWARE (SAS)	15
2.5.3	MINITAB16	16
2.5.4	STATA	16
2.6	SOFTWARE: PAQUETES COMERCIALES	16
2.6.1	GGE-BILOT	17
2.6.2	MULTIVARIATE STATISTICAL PACKAGE (MVSP)	17
2.6.3	STATGRAPHICS	18
2.6.4	PC-ORD	18
2.6.5	CANOCO	19
2.6.6	ANALYSE-IT	19
2.7	SOFTWARE: PAQUETES LIBRES	20
2.7.1	XLS-BILOT	20
2.7.2	VISTA THE VISUAL STATISTICS SYSTEM	21
2.7.3	BRODGAR	21
2.7.4	MULTBILOT: MULTIVARIATE ANALYSIS USING BILOTS	22
2.8	PAQUETES DE R	23
2.8.1	STATS	23

2.8.2	MULTIBILOTGUI	24
2.8.3	BILOTBOOTGUI.....	24
2.8.4	GGEBILOT	24
2.8.5	ADE4.....	25
2.8.6	VEGAN	25
2.8.7	BILOTML.....	26
2.8.8	MULTBILOTR	26
2.9	SISTEMA DE INFORMACIÓN HOSPITALARIO (HIS).....	27
2.10	SISTEMA DE INFORMACIÓN DE LABORATORIO CLÍNICO (LIS).....	28
2.11	LENGUAJES DE PROGRAMACIÓN WEB DE CÓDIGO ABIERTO	30
2.11.1	PHP.....	30
2.11.2	JAVA.....	31
2.11.3	PYTHON.....	32
2.12	HERRAMIENTAS PARA DISEÑO WEB DE CÓDIGO ABIERTO	32
2.12.1	NETBEANS	33
2.12.2	ECLIPSE.....	33
2.12.3	WORDPRESS.....	34
2.12.4	BOOTSTRAP.....	35
2.13	BIBLIOTECAS DE SOFTWARE GRÁFICAS PARA APLICACIONES WEB DE CÓDIGO ABIERTO	36
2.13.1	FUSIONCHARTS	36
2.13.2	CHART.JS	37
2.13.3	HIGHCHARTS	37
2.14	BIBLIOTECAS DE SOFTWARE INTERACTIVAS PARA APLICACIONES WEB DE CÓDIGO ABIERTO.....	38

2.14.1	JAVASCRIPT	38
2.14.2	JQUERY	39
CAPÍTULO 3		41
3.	METODOLOGÍA	41
3.1	INTRODUCCIÓN	41
3.2	ANÁLISIS DE LA APLICACIÓN WEB INTERACTIVA	43
3.2.1	LENGUAJE DE PROGRAMACIÓN WEB	43
3.2.2	HERRAMIENTAS DE VISUALIZACIÓN DE APLICACIONES WEB	44
3.2.2.1	HERRAMIENTAS PARA DISEÑO DE APLICACIONES WEB	45
3.2.2.2	BIBLIOTECAS DE SOFTWARE INTERACTIVAS PARA APLICACIONES WEB	47
3.2.3	FUENTES DE DATOS	48
3.3	DISEÑO DE LA APLICACIÓN WEB INTERACTIVA	49
3.3.1	PREPARACIÓN DEL CONJUNTO DE DATOS	49
3.3.2	DISEÑO DE OPCIONES Y FUNCIONALIDADES DE APLICACIÓN WEB INTERACTIVA	52
3.3.2.1	SELECCIÓN DE ARCHIVO	53
3.3.2.2	ARCHIVO SELECCIONADO	53
3.3.2.3	CONFIGURACIONES GENERALES	54
3.3.2.4	CONFIGURACIONES PARA GRAFICAS DE LÍNEAS Y CIRCULAR	56
3.3.2.5	CONFIGURACIONES PARA MATRIZ DE CORRELACIÓN	57
3.3.2.6	CONFIGURACIONES PARA DIAGRAMA DE CAJAS	57
3.3.2.7	CONFIGURACIONES PARA HJ BILOT	58
3.3.2.8	CONFIGURACIONES PARA CLÚSTER EN HJ BILOT	62
3.3.2.9	CONFIGURACIONES PARA SPARSE BILOT	62

3.4	DESARROLLO DE LA APLICACIÓN WEB INTERACTIVA.....	64
3.4.1	INTERFAZ WEB DE VISUALIZACIÓN	65
3.4.2	GESTIÓN DE ARCHIVOS.....	65
3.4.3	GRÁFICOS INTERACTIVOS Y DINÁMICOS	65
3.4.4	GRÁFICOS PARA VISUALIZACIÓN EXPLORATORIA.....	66
3.4.5	GRÁFICOS PARA VISUALIZACIÓN AVANZADA	67
CAPÍTULO 4		70
4.	RESULTADOS	70
4.1	INTRODUCCIÓN.....	70
4.2	VISUALIZACIÓN EXPLORATORIA INTERACTIVA.....	70
4.2.1	GRAFICOS DE LÍNEAS.....	70
4.2.2	GRAFICOS CIRCULARES.....	71
4.2.3	DIAGRAMA DE CAJAS	72
4.3	VISUALIZACIÓN AVANZADA INTERACTIVA	73
4.3.1	HJ BILOT	73
4.3.2	CLÚSTER EN HJ BILOT.....	83
4.3.3	ANÁLISIS DE COMPONENTES PRINCIPALES SPARSE (SPCA).....	89
CAPÍTULO 5		95
5.	CONCLUSIONES Y RECOMENDACIONES	95
6.	REFERENCIAS.....	97
7.	GLOSARIO.....	100
8.	APÉNDICES Y ANEXOS.....	101

ABREVIATURAS O SIGLAS

AJAX	Asynchronous JavaScript and XML
CRAN	Comprehensive R Archive Network
CSS	Cascading Style Sheets
CSV	Comma Separated Values
CVS	Concurrent Versions System
DOM	Document Object Model
DVS	Descomposición en Valores Singulares
GPL	General Public License
HTML	HyperText Markup Language
IDE	Integrated Development Environment
IESS	Instituto Ecuatoriano de Seguridad Social
ISSFA	Instituto de Seguridad Social de la Fuerzas Armadas del Ecuador
ISSPOL	Instituto de Seguridad Social de la Policia Nacional del Ecuador
JDK	Java Development Kit
JSON	JavaScript Object Notation
LIS	Laboratory Information System
MSP	Ministerio de Salud Pública del Ecuador
PCA	Principal Component Analysis
PHP	Preprocesador de Hipertexto
RAM	Random Access Memory
SPSS	Statistical Product and Service Solutions

SQL Structured Query Language

XML Extensible Markup Language

LISTADO DE FIGURAS

Figura 1.1. Representación HJ Biplot.....	3
Figura 2.1. Representación del Disjoint Biplot.....	12
Figura 3.1. Metodología.	43
Figura 3.2. Conjunto de Datos.	52
Figura 3.3. Selección de Archivo.....	53
Figura 3.4. Archivo seleccionado.	54
Figura 3.5. Configuraciones Generales.....	55
Figura 3.6. Configuraciones para Gráficos de Líneas y Circular.	57
Figura 3.7. Configuraciones para Matriz de Correlación.	57
Figura 3.8. Configuraciones para Diagrama de Cajas.	58
Figura 3.9. Configuraciones para HJ Biplot.....	60
Figura 3.10. Funcionalidades para HJ Biplot.....	62
Figura 3.11. Configuraciones de Clusters para HJ Biplot.....	62
Figura 3.12. Configuraciones para Sparse Biplot.....	64
Figura 3.13. Gráfico de Línea Interactivo y Dinámico.	66
Figura 3.14. Gráfico Circular Exploratorio Interactivo y Dinámico.	67
Figura 3.15. HJ Biplot Interactivo y Dinámico.....	68
Figura 4.1. Gráfico de Línea Interactivo y Dinámico.	71
Figura 4.2. Gráfico Circular Interactivo y Dinámico.	72
Figura 4.3. Diagrama de Cajas Interactivo y Dinámico.	73
Figura 4.4. HJ Biplot Interactivo y Dinámico en 3D.	77
Figura 4.5. HJ Biplot Interactivo y Dinámico en 2D.	82
Figura 4.6. Contribución de Variables sobre las Dimensiones.....	83
Figura 4.7. Clúster por Lugar de Nacimiento en HJ Biplot Interactivo y Dinámico.	87
Figura 4.8. Clúster por Diagnóstico en HJ Biplot Interactivo y Dinámico.	89
Figura 4.9. Matriz de Correlación de Variables.	89
Figura 4.10. Porcentaje de varianza geométrica explicada (ACP Sparse).....	91
Figura 4.11. Análisis de Componentes Principales Sparse (SPCA).....	92

Figura 4.12. Porcentaje de varianza geométrica explicada (ACP Sparse) con penalización Lasso (λ).....	93
Figura 4.13. Análisis de Componentes Principales Sparse (SPCA) con penalización Lasso (λ).	94
Figura A1. Código de Programa PHP para Interfaz Web de Visualización.	101
Figura A2. Código de programa PHP para Gráfico Interactivo y Dinámico.	102
Figura A3. Código de programa PHP para Gestión de Archivos.....	103
Figura A4. Código de programa R para Gráfico Exploratorio Circular.	104
Figura A5. Código de programa R para Sparse PCA.....	105

LISTADO DE TABLAS

Tabla 1. Matriz de Correlaciones de Variables.....	90
--	----

CAPÍTULO 1

1. INTRODUCCIÓN

El presente trabajo servirá como una herramienta de visualización interactiva que permitirá emplear técnicas de Data Mining para el análisis de los datos, que conduzcan a mejorar su eficiencia, seguridad y calidad en el procesamiento de la información, representándolos gráficamente por medio de los métodos de la familia Biplot, haciendo uso de herramientas web de código abierto.

En el mismo sentido, se describen los lenguajes de programación web, y las herramientas que permitan implementar Aplicaciones Web Interactivas, como aporte para la representación gráfica de los métodos de la familia Biplot.

1.1 ANTECEDENTES

HJ-Biplot (Galindo, 1986), es una técnica de análisis multivariante que permite representar tanto individuos como variables en un espacio de dimensiones reducidas; este método, emerge como herramientas de visualización poderosas y confiables similares a un diagrama de dispersión pero que capturan las estructuras de covarianza de datos multivariante. La versión inferencial del HJ-Biplot puede consultarse en (Nieto et al., 2014).

Para adaptar este enfoque a conjuntos de datos masivos, es necesario implementar nuevas técnicas que sean capaces de reducir la dimensionalidad de

los datos y mejorar la interpretación. En este trabajo, se programará computacionalmente algunos métodos Biplot (Gabriel, 1971): como, por ejemplo, HJ-Biplot clásico, Clustering Disjoint HJ-Biplot (CDBiplot) (Nieto et al., 2017), SparseBiplots (Cubilla, 2021) y HJ-Biplot para datos composicionales (Hernández, 2016). Todas estas técnicas serán integradas en la Herramienta Interactiva Web que es el núcleo central del trabajo.

En la actualidad existe una gran cantidad de información almacenada en diversos gestores de bases de datos, sin embargo; usar las herramientas convencionales para la gestión de la información contenida en las fuentes de datos es limitada, debido a que generalmente no incluyen un proceso de análisis de los datos y extracción para la generación del nuevo conocimiento (González, 2023).

1.2 DESCRIPCIÓN DEL PROBLEMA

En este tiempo disponemos de volúmenes de información considerables almacenados en numerosas bases de datos, no obstante; trabajar con los sistemas convencionales para el manejo de información contenida en dichas fuentes tiene sus limitaciones ya que la mayoría de los sistemas no incluye un proceso de análisis de la información y extracción que permita incorporar nuevo conocimiento. Hoy en día existen nuevas técnicas y disciplinas para extraer información útil a partir de distintas fuentes de información. Esto permite tomar decisiones importantes sobre un tema en particular. Estas técnicas han sido implementadas en herramientas construidas con el objetivo de aplicar la minería de datos o el proceso de descubrimiento de conocimiento.

El almacenamiento de los datos que existe en nuestros días ha supuesto un gran reto para las herramientas de visualización interactivas. Por tanto, es esencial para las áreas asistenciales de los Hospitales Oncológicos, comprender las relaciones de manera visual, entre la historia clínica del paciente, su estadificación y la patología diagnóstica del cáncer, mediante técnicas de minería de datos empleando métodos de la familia Biplot, como se ilustra en la Figura 1.1.

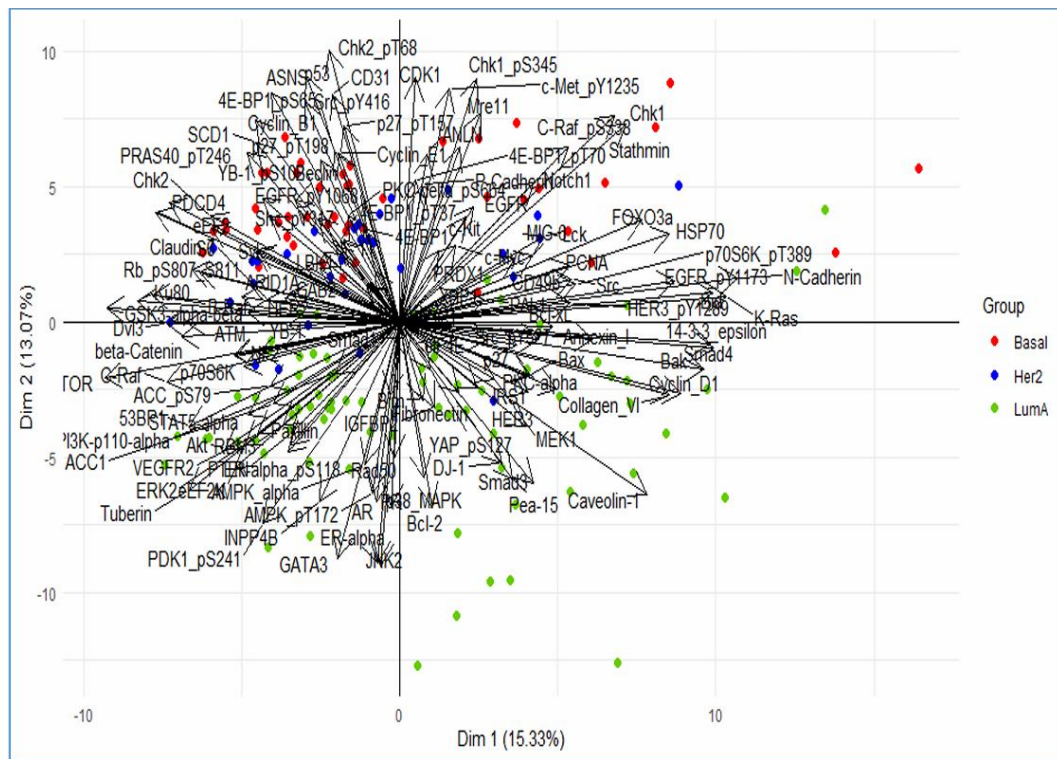


Figura 1.1. Representación HJ Biplot

Nota. El gráfico representa un patrón de luz entre tres tipos de cáncer y las proteínas, pero no está del todo claro, sin embargo; debido al elevado número de proteínas, es difícil identificar la contribución en cada eje (Cubilla et al., 2021).

1.3 SOLUCIÓN PROPUESTA

En el Ecuador no existe evidencia científica sobre métodos estadísticos que permitan conocer la relación que existe entre las pruebas de laboratorio clínico y la estadificación del cáncer en el paciente; consecuentemente nuestro estudio ayudará a generar nuevos conocimientos sobre el estado del cáncer para que el personal asistencial pueda elaborar tratamientos alternativos que aporten a la prevención del padecimiento de esta enfermedad catastrófica.

Basado en las herramientas de programación actuales se propondrá analizar, diseñar, desarrollar e implementar una Aplicación Web Interactiva para un proceso de Data Mining basado en métodos de la familia Biplot, que permita analizar la relación entre pruebas de laboratorio clínico y la estadificación en pacientes con patología de cáncer.

Para la implementación de la Aplicación Web Interactiva, será necesario el uso de: lenguajes de programación web de código abierto, herramientas de diseño de código abierto para aplicaciones web, bibliotecas de software gráficas de código abierto para aplicaciones web, bibliotecas de software interactivas de código abierto para aplicaciones web.

Además, con el desarrollo de esta propuesta, la Aplicación Web Interactiva permitirá integrar las técnicas de Data Mining con los métodos de la familia Biplot: HJ-Biplot Clásico, Clustering HJ-Biplot, y SparseBiplots.

1.4 OBJETIVO GENERAL

Integrar de manera interactiva los métodos de Biplot en una Aplicación Web, mediante lenguajes de programación de código abierto y probar su funcionalidad con datos reales de un Hospital de Pacientes Oncológicos.

1.5 OBJETIVOS ESPECÍFICOS

- Revisar la bibliografía científica para identificar las alternativas a los métodos de Biplots clásicos propuestos para trabajar con datos masivos.
- Diseñar una Interfaz Web intuitiva y funcional que facilite a los usuarios la visualización de indicadores clave y resultados de los análisis multivariantes Biplot.
- Programar, en lenguaje SQL, la estrategia para la obtención automática del conjunto de variables de interés, a partir de un Sistema de Gestión de Base de Datos masivos de Pacientes Oncológicos.
- Analizar la relación entre las pruebas de laboratorio clínico, y la estadificación del cáncer, con la ayuda de la Aplicación Web diseñada en esta investigación, para la detección de posibles patrones que desencadenen la enfermedad.

1.6 ALCANCE

Este proyecto se focalizará en la ciudad de Guayaquil, en el periodo comprendido entre 2019-2023 y su población accesible será todos los pacientes que fueron atendidos en la unidad de preadmisión de un hospital oncológico a través de la plataforma Intranet de un Sistema de Información Hospitalario (HIS) en pacientes

con financiamiento propio o financiados tanto por la Red Pública Integral de Salud (IESS, ISSFA, ISSPOL y MSP), como por la Red complementaria.

CAPÍTULO 2

2. MARCO TEÓRICO

2.1 BIPLOTS CLASICOS

Un biplot (González, 2023) es una representación gráfica de datos multivariantes. De la misma manera que un diagrama de dispersión muestra la distribución conjunta de dos variables, un biplot representa tres o más (Gabriel y Odoroff, 1990).

El biplot (González, 2023) aproxima la distribución de una muestra multivariante en un espacio de dimensión reducida, normalmente de dimensión dos, y superpone sobre la misma representación de las variables sobre las que se mide la muestra (Gower y Hand, 1995). Las representaciones de las variables son normalmente vectores, y coinciden con las direcciones en las que mejor se muestra el cambio individual de cada variable.

Como ya habíamos mencionado en la introducción, el prefijo "bi" se refiere a la superposición, en la misma representación, de individuos y variables.

De acuerdo con Gabriel (1971), las representaciones biplot se basan en la reducción de la dimensión a través de la factorización de la matriz de partida en el producto escalar de otras dos.

Si X es la matriz de datos (transformados adecuadamente) que queremos representar, con I filas y J columnas, la factorización puede escribirse como:

$$X = AB^T + E \quad (2.1)$$

donde las filas de A definen un conjunto de puntos que usaremos como marcadores fila, las filas de B como marcadores columna y E la matriz de errores o residuales.

La versión original utiliza como factorización la denominada *Descomposición en Valores Singulares (DVS)* estrechamente relacionada con el Análisis de Componentes Principales y el Análisis Factorial. La DVS puede definirse como:

$$X = UAV^T = \sum \lambda_r u_r v_r^T \quad (2.2)$$

donde u_r son los vectores singulares por la derecha contenidos en V , es decir, los vectores propios de $X^T X$; u_r son los vectores singulares por la izquierda recogidos en la matriz U , es decir, los vectores propios de XX^T , λ_r los valores singulares no negativos ordenados de forma decreciente. Los cuadrados de los valores singulares λ_r^2 son también los valores propios no nulos de $X^T X$ y XX^T , que coinciden.

Para la matriz X cuyo rango es $R \leq \min(I, J)$, es posible obtener una aproximación (\hat{X}) de bajo rango ($S < R$) tomando los primeros S términos de la DVS.

$$X = \hat{X} + E = U_{(S)} \Lambda_{(S)} V_S^T + U_{(-S)} \Lambda_{(-S)} V_{-S}^T = \sum_{r=1}^S \lambda_r u_r v_r^T + \sum_{r=S+1}^R \lambda_r u_r v_r^T \quad (2.3)$$

siendo (S) las S primeras columnas y $(-S)$ el resto de las columnas.

Es posible representar esta factorización con un biplot \hat{X} , en dos o tres dimensiones, si puede considerarse que tiene una bondad de ajuste adecuada. Para calcular la bondad de ajuste de la aproximación a bajo rango se calcula el cociente entre la suma de cuadrados de los primeros S valores singulares y la suma de cuadrados de todos. Generalmente se presenta mediante porcentajes.

$$\frac{\sum_{r=1}^S \lambda_r^2}{\sum_{r=1}^R \lambda_r^2} * 100 \quad (2.4)$$

Volviendo a la factorización inicial (ecuación 2.1), una forma general para el biplot es tomar como marcadores fila a la matriz $A = U\Lambda^\gamma$ y como marcadores columna a $B = V\Lambda^{1-\gamma}$ con γ comprendida entre 0 y 1.

En función del valor γ seleccionado se obtienen los diferentes tipos de biplot clásicos:

2.1.1 JK-BILOT

También denominado RMP-Biplot, con $\gamma = 1$. Los marcadores de las filas pueden definirse como $J_{(S)} = U_{(S)} \Lambda_{(S)}$ y coinciden, en el espacio de las componentes principales, con las coordenadas de los individuos. Los marcadores columna serán las proyecciones de los ejes originales en

este mismo espacio y se pueden definir como: $K_{(s)} = V_{(s)}$. La calidad de representación en este tipo de gráfico es mejor en filas que en columnas.

2.1.2 GH-BILOT

Con $\gamma = 0$, y también denominado CMP-Biplot. Los marcadores de las filas pueden definirse como $G_{(s)} = \sqrt{I-1} U_{(s)}$, cuya distancia se aproxima a la distancia de Mahalanobis en el espacio multidimensional con baja calidad de representación. Las coordenadas de las columnas $H_{(s)} = \frac{1}{\sqrt{I-1}} \Lambda_{(s)} V_{(s)}$ son las cargas de un modelo de Análisis Factorial si los datos están estandarizados. Las varianzas y covarianzas entre las variables se aproximan de forma que las correlaciones entre las variables correspondan con los cosenos de los ángulos entre ellas y la variabilidad a través de la longitud del vector.

2.1.3 SQRT-BILOT

Cuando $\gamma = 1/2$ el biplot obtenido no está relacionado con las técnicas más conocidas. Los marcadores fila son $U_{(s)} \lambda_{(s)}^{1/2}$ y los marcadores columna como $V_{(s)} \lambda_{(s)}^{1/2}$. La calidad de representación de las entradas de la matriz de datos se mantiene.

2.2 HJ-BILOT

Los Biplots clásicos permiten la representación simultánea de las filas y las columnas de una matriz, sin embargo, la calidad de representación de las y columnas no es la misma ya que cambia en función del tipo de biplot que se esté utilizando. Si se busca que la representación simultánea mantenga la calidad de individuos y variables, Galindo Villardón (1986) propone el HJ-Biplot, que es una representación simétrica tal y como se define en el Análisis de Correspondencias.

Para una matriz de datos X , un HJ-Biplot es una representación gráfica multivariante cuyos marcadores la serán identificados con la matriz J y los marcadores columna por H .

Partiendo de la descomposición en valores y vectores propios de la matriz X como en la ecuación 3.2, los marcadores se definen de la siguiente forma:

$$J_{(S)} = U_{(S)} \Lambda_{(S)}$$

$$H_{(S)} = V_{(S)} \Lambda_{(S)}$$

Tanto los marcadores fila como los marcadores columna comparten los valores propios, se realizará la representación sobre el mismo sistema de referencia.

2.3 CDBILOT

Definido como Disjoint Biplot (DBiplot), es un método que construye ejes factoriales disjuntos garantizando que cada variable de la matriz de datos original contribuya solamente a una componente principal, como se ilustra en la Figura 2.1. El algoritmo parte de una clasificación aleatoria de las variables en las

componentes principales y mediante un procedimiento iterativo busca la clasificación óptima que conduzca a la maximización de la variabilidad explicada. La representación gráfica de objetos y variables en este nuevo espacio de dimensión reducida se realiza a través del HJ-Biplot. Para ello, una función denominada CDBiplot se aloja dentro de la interfaz gráfica biplotbootGUI (Nieto-Librero & Galindo-Villardón, 2015). La interfaz posee tres funciones principales. La función CDBiplot ejecuta la interfaz gráfica para construir el Disjoint Biplot (DBiplot); la representación de clústers, mediante la función Clustering Biplot (CBiplot), y el Clustering Disjoint Biplot permite la representación conjunta tanto de individuos como de variables.

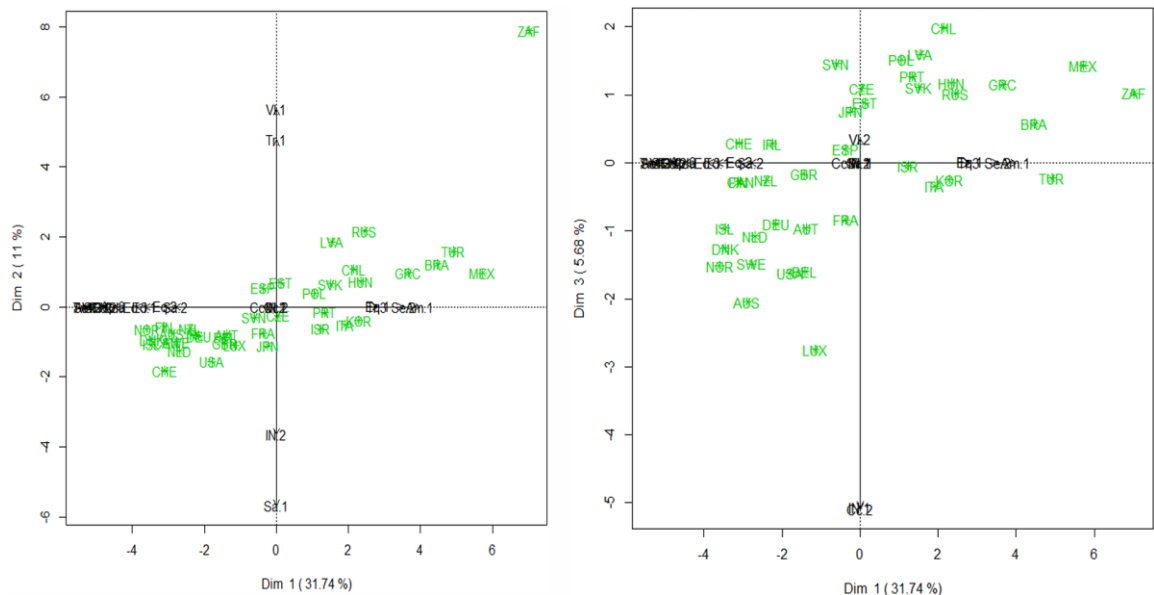


Figura 2.1. Representación del Disjoint Biplot.

Nota: Gráfico izquierda (plano 1-2). Gráfico derecho (plano 1-3).

2.4 SPARSE BIPLLOT

En el contexto del Biplot, no hemos encontrado ninguna evidencia que formule algoritmos alternativos para penalizar o contraer las cargas de las componentes principales, con el fin de mejorar la interpretación de la información que aportan los datos de alta dimensionalidad. Se propone nuevas alternativas de representación Biplot que consisten en adaptar restricciones para contraer y/o producir cargas nulas en las componentes, en base a las teorías de regularización Ridge, LASSO y Elastic Net (Cubilla, 2019). En cada caso se demuestra el funcionamiento de los algoritmos mediante la creación en lenguaje R del paquete “SparseBiplots”, diseñado exclusivamente para dar soporte a la nueva metodología planteada. La implementación del paquete se realiza con los datos del Índice para una Vida Mejor. Además, se utiliza una muestra en la que se analiza los indicadores sociales de grandes compañías que informan sobre Responsabilidad Social Corporativa (RSC) de conformidad con el modelo del Global Reporting Initiative (Cubilla-Montilla, Nieto-Librero, Galindo Villardón, Vicente Galindo, & Garcia-Sanchez, 2019).

El Biplot Sparse aborda el problema de encontrar una combinación lineal de las variables, determinado por un vector de cargas sparse que maximiza la variabilidad de los datos o minimiza el error de construcción. Este enfoque mejora notablemente la capacidad de interpretar los ejes (Sparse) obtenidos.

El Biplot se reformula como un modelo de regresión lineal simple, bajo el concepto de minimización del “Error de reconstrucción (E)”.

$$E = \left[\|X - \hat{X}\|^2 \right] = TR \left(E \left[(X - \hat{X})(X - \hat{X})^T \right] \right)$$

Esto quiere decir que se busca minimizar la diferencia entre la matriz original y los datos que se obtendrían proyectando en el espacio original las p nuevas variables.

La formulación del Biplot como un problema de regresión, impone restricciones en las cargas factoriales para producir “ejes modificados”. Obviamente, la incorporación de una restricción adicional proporciona unas dimensiones que, en general, no explican toda la varianza que se explica en las dimensiones originales. No obstante, las técnicas de regularización, proporciona cierta estabilidad en el proceso y mejora la capacidad de generalización del modelo. El parámetro de regularización λ controla la fuerza o importancia que le damos a la regularización en el proceso de optimización.

2.5 SOFTWARE BILOT

Para que las técnicas descritas a lo largo del capítulo sean útiles, es necesario que exista al menos un software que permita realizar tanto los cálculos como las representaciones gráficas. El biplot es una técnica que, debido a su gran utilidad, su uso es cada vez mayor y por ello existen diversos paquetes, programas y aplicaciones que han incluido una sección dedicada al biplot o que se han desarrollado específicamente para su uso. En esta subsección recogeremos algunos de los más importantes y que están siendo más utilizados en la actualidad.

Dentro de los softwares de uso general que están ampliamente extendidos, tenemos: SPSS, SAS Minitab16 o STATA, que contienen un apartado específico para realizar este tipo de técnicas, aunque en todos los casos contiene únicamente los Biplots clásicos.

2.5.1 IBM SPSS STATISTICS

Para realizar un Análisis Biplot dentro de este software será necesario realizar en primer lugar el análisis que denominan "Análisis de componentes principales categórico (CATPCA)". Este análisis se encontrará dentro del menú de *Reducción de dimensiones, Escalamiento óptimo*; dentro, deberá ser elegido el biplot en el apartado de gráficos, el *diagrama de dispersión biespacial* (IBM Corp., 2024).

2.5.2 STATISTICAL ANALYSIS SOFTWARE (SAS)

Este programa contiene varias opciones para la realización de cuatro tipos de biplot, los tres biplots clásicos y el biplot de covarianza (biplot-COV). Las opciones que presenta este software (SAS Institute Inc., 2024) son las siguientes:

- Por medio del menú PROC PRINQUAL del software SAS/STAT para crear el biplot COV.
- Con la licencia para el software SAS/GRAPH y para el software SAS/IML es posible usar la macro %BIPLOT de Michael Friendly

empleando la opción OUT = en ella. Es posible crear una versión más moderna empleando PROC SGPLOT.

- Realizar los cálculos matriciales con SAS/IML para obtener las coordenadas de los marcadores y vectores. A continuación, se empleará el módulo de biplot para realizar la representación gráfica.
- Utilizar el módulo WriteBiplot para realizar los cálculos y PROC SGPLOT para crear el biplot.

2.5.3 MINITAB16

En el caso de Minitab, igual que ocurría en SPSS, debemos realizar un Análisis de Componentes Principales, y a continuación, en el menú de gráficos que se encuentra dentro del análisis anterior, será necesario seleccionar la gráfica de doble proyección (LLC, 2024).

2.5.4 STATA

En el software STATA la función biplot será utilizada para hacer este tipo de representaciones gráficas (StataCorp, 2024). La función se encuentra en el menú de estadísticos, dentro del apartado de estadística multivariante.

2.6 SOFTWARE: PAQUETES COMERCIALES

Debido a que el software de uso general no permite realizar un gran número de tipos de biplots, se han creado recientemente otros paquetes comerciales que amplían el número de representaciones gráficas que se pueden realizar.

Algunas de estas son GGE-Biplot, MVSP, Statgraphics, PC-ORD, CANOCO y Analyse-it, descritos a continuación:

2.6.1 GGE-BILOT

En este software, además de los biplots clásicos, se pueden construir otros biplots como el AMMI biplot, el GGE biplot (genotipo-ambiente), biplot de expresión génica, biplot QTL de mapeo, etc.

Este software está orientado al ámbito de la Agronomía, pero puede ser utilizado en cualquier ámbito (Yan y Kang, 2006).

2.6.2 MULTIVARIATE STATISTICAL PACKAGE (MVSP)

Este software se especializa en la realización de Componentes Principales, Coordenadas Principales y Análisis de Correspondencias (Kovach, 1999). Dentro de sus funciones se encuentra la posibilidad de realizar representaciones biplot con los resultados obtenidos.

Para elaborar una representación biplot en este programa será necesario representarla tras haber obtenido los cálculos del PCA o del CCA.

El campo que utiliza este software por excelencia es el ámbito de la Ecología.

2.6.3 STATGRAPHICS

Este programa ha realizado recientemente nuevas incorporaciones, entre ellas la conexión con R y con Python desde su interfaz gráfica (Inc. Statgraphics Technologies, 2024).

Incorpora un gran número de técnicas univariantes y multivariantes entre sus menús.

Para construir el biplot en este software debemos seleccionar, dentro del menú Avanzado, los Métodos Multivariantes, y a continuación Componentes Principales y las opciones avanzadas de los gráficos.

Este software dará la posibilidad de realizar gráficos en 3D.

2.6.4 PC-ORD

Una de las características de este software es que solo se ha desarrollado para el sistema operativo Windows (James Grace y Hatch, 2018).

Es un paquete muy amplio que contiene multitud de técnicas multivariantes, sobre todo las más utilizadas dentro del campo de la Ecología. Entre las funciones disponibles dentro del paquete se

encuentra un menú para realizar diferentes tipos de Biplots. Este software también incluye la posibilidad de realizar GGE biplots.

2.6.5 CANOCO

Este programa solo se encuentra disponible para el sistema operativo Windows (Šmilauer, 2012).

CANOCO fue desarrollado para realizar el Análisis Canónico de Correspondencias, sin embargo, en la actualidad contiene un gran número de análisis multivariantes que pueden realizarse, incluyendo la realización de Biplots dentro de su ventana de visualización de los datos.

El campo en el que más se utiliza este software es el ámbito de la Ecología.

2.6.6 ANALYSE-IT

El Software Analyse-it está implementado como un complemento de Excel, sin embargo, no está disponible para las versiones de Excel de MacOs (Ltd. Analyse-it Software, 2022).

Este complemento aporta a las hojas de cálculo de Excel la posibilidad de realizar técnicas estadísticas más complejas, construcción de modelos, control de calidad o métodos de validación entre otras.

Las técnicas de mayor interés en este capítulo son las de Análisis Multivariante, concretamente las relacionadas con la realización de Análisis de Componentes Principales, ya que dentro de este menú se puede encontrar un menú que permite la representación biplot de las variables de estudio.

2.7 SOFTWARE: PAQUETES LIBRES

Los paquetes comerciales, debido al costo, no siempre son las opciones óptimas para crear los Biplots. Además, en muchos casos, están cerrados a la incorporación de nuevos métodos, consecuentemente, los Biplots de reciente creación no forman parte de los paquetes.

De la misma forma que se crean los paquetes comerciales, se han ido desarrollando paquetes de código libre que contienen más alternativas, desde complementos de Excel, software propio o paquetes de R.

En este apartado describiremos los paquetes disponibles de forma libre: XLS-Biplot, ViSta, Brodgar y MultBiplot.

2.7.1 XLS-BILOT

Se trata de un complemento para Excel en el que se incluye la construcción de este tipo de representaciones gráficas (Udina, 2005).

Para utilizarla será necesario instalar la macro correspondiente y abrir, dentro de los Complementos, la opción Biplot. A continuación,

deberemos elegir la *Descomposición en Valores Singulares*, seleccionar los datos de análisis y sus correspondientes etiquetas, y elegir el método de cálculo y las transformaciones. Por último, aparecerá la ventana para la realización del Biplot que contiene las características del gráfico y, en el escalado, es posible elegir entre los Biplots clásicos para realizar la representación.

Este software fue descrito en un artículo por Udina (2004) donde se puede encontrar más información.

2.7.2 VISTA THE VISUAL STATISTICS SYSTEM

Este software ha sido desarrollado como una forma de representar gráficamente datos creando visualizaciones dinámicas (Young, 1990).

No se ha actualizado desde hace tiempo, sin embargo, contiene un apartado que permite realizar Biplots dentro de las representaciones gráficas de las Componentes Principales de los Análisis Multivariantes creados en el Software.

2.7.3 BRODGAR

Para el uso de este software será necesario tener instalado R ya que lo usará como compilador. Son programas separados, aunque existe una versión para R. Las funciones en ambos casos son mayoritariamente las

mismas, sin embargo, hay algunas más dentro del software externo (Highland Statistics Ltd., 2017).

No existen actualizaciones recientes y será un software libre hasta el año 2026, teniendo que renovar el código de licencia de forma anual.

Este paquete contiene un gran número de análisis, tanto univariantes como multivariantes, que se pueden realizar a través de su interfaz gráfica.

Por defecto, Brodgar realizará un Biplot de correlaciones, aunque es posible seleccionar el tipo de biplot que se desea representar.

2.7.4 MULTBILOT: MULTIVARIATE ANALYSIS USING BILOTS

Este software está asociado a Matlab (Vicente-Villardón, 2024), sin embargo, no es necesario haber descargado Matlab para poder operar con él, si se realiza la descarga completa. También existe una versión dentro de R y una versión integrada dentro de Matlab.

El paquete descargado, en este caso, contiene código menos actualizado y un menor número de técnicas que el paquete de R. Su interfaz gráfica está disponible tanto para el sistema operativa Windows como para MacOs.

Como en otros de los paquetes presentados hasta el momento, el paquete MultiBiplot contiene un gran número de técnicas multivariantes y

todas ellas presentadas en forma de Biplot, entre ellas es posible destacar el Análisis de Componentes Principales, el Análisis Factorial, el Unfolding o el STATIS-ACT.

De los programas presentados, este es el único que contiene los Biplots logísticos.

2.8 PAQUETES DE R

R es un software donde podemos realizar un gran número de Biplots (R Core Team, 2021), será el que emplearemos para el Data Mining y la generación de código en formato JSON que permita ilustrar los gráficos por medio de una Aplicación Web Interactiva.

Al tratarse de un software libre, se encuentran multitud de paquetes que permiten realizar este tipo de representaciones gráficas. Los autores de cada paquete incluyen los Biplots que son de interés con las técnicas que se han incluido en el mismo.

La mayor parte de los paquetes presentados a continuación es posible encontrarlos en el repositorio CRAN.

2.8.1 STATS

El paquete de estadística básico de R contiene una función ("biplot") que permite realizar los biplots clásicos sobre los resultados de unas Componentes Principales.

2.8.2 MULTIBILOTGUI

Este paquete contiene una interfaz gráfica que permite realizar un biplot asociado al Análisis Factorial Múltiple (Librero et al., 2022).

Es posible elegir el tipo de biplot utilizado entre los biplots clásicos y el HJ-biplot.

2.8.3 BILOTBOOTGUI

En el paquete biplotbootGUI (Librero et al., 2019) las autoras introducen métodos bootstrap a los biplots clásicos para obtener las regiones de confianza a través de una interfaz gráfica de R.

Con esta interfaz también es posible realizar los cálculos del Clustering Disjoint Biplot, así como el Análisis de Componentes Principales asociado a este tipo de biplot.

2.8.4 GGEBILOT

Este paquete contiene las funciones necesarias para construir y evaluar un biplot genotipo ambiente (Dumble, 2022). Es posible realizar la comparación de dos genotipos en todos los ambientes sometidos a estudio, así como establecer la relación entre los ambientes, examinar por separado genotipos y ambientes, o construir un orden de cada uno de ellos.

Recientemente ha sido eliminada de CRAN la interfaz gráfica de este paquete por compatibilidad con las versiones más recientes.

2.8.5 ADE4

El paquete desarrollado por el Laboratorio de Biometría y Biología Evolutiva (UMR CNRS5558) de la Universidad de Lyon contiene un conjunto de funciones que permiten analizar datos Ecológicos y Ambientales utilizando métodos de tres vías (Dray et al., 2022).

Entre la multitud de funciones que se pueden encontrar en este paquete, encontramos diversas representaciones gráficas, una de ellas son los biplots presentados en este capítulo. Utilizando la función "scatter" haremos una representación básica. Existen funciones dentro del paquete para realizar biplots asociados a los mapas factoriales de diferentes Análisis de Correspondencias, a un PLS con el algoritmo NIPALS o a un Análisis de Coordenadas Principales.

2.8.6 VEGAN

Este paquete está diseñado para los análisis ecológicos. Contiene un gran número de técnicas de gran utilidad en este campo (Oksanen et al., 2017).

Este paquete contiene, varias funciones que permiten realizar representaciones biplot. Estos biplots estarán asociados

fundamentalmente al Análisis Canónico de Correspondencias y al Análisis de la Redundancia.

2.8.7 BILOTML

El paquete BiplotML, de reciente creación, será utilizado para realizar representaciones de biplots logísticos (Babativa-Márquez, 2020).

Aunque en este documento se ha descrito la forma tradicional de calcular los parámetros, este paquete implementa nuevas metodologías que optimizan la estimación de los parámetros del modelo y construye los Biplot Logísticos a partir de ellos.

2.8.8 MULTBILOTR

Este paquete es una versión mejorada de la interfaz gráfica desarrollada en Matlab y contiene un mayor número de recursos que el paquete libre de partida (Vicente-Villardón, 2021).

Este paquete contiene varias técnicas multivariantes desde una perspectiva de biplot. Son muchos los análisis que se incluyen en este paquete, desde innovaciones hasta técnicas clásicas. Algunos de ellos son los Biplots clásicos, HJ-Biplot, Biplots canónicos, MANOVA Biplots, Análisis de Correspondencia, Análisis de Correspondencia Canónico, STATIS-ACT canónico, Biplots logísticos para datos binarios y ordinales, Despliegue multidimensional, Biplots externos para el Análisis de

Coordenadas Principales o el Escalado Multidimensional, entre muchos otros.

En los siguientes apartados, se exponen las principales teorías y conceptos referente a los sistemas de información, lenguajes de programación web, herramientas para diseño de aplicaciones web, y bibliotecas de software para el desarrollo de aplicaciones web.

2.9 SISTEMA DE INFORMACIÓN HOSPITALARIO (HIS)

Hospital Information System (HIS), sistema de información hospitalario por sus siglas en inglés, también conocido como sistema de información de salud, es un sistema de información que permite, generar, procesar, modificar, recuperar y almacenar información así como también reinterpretar datos, emitir reportes e informes de tipo médico-administrativos de una institución hospitalaria, optimizando los recursos humanos y materiales, para ofrecer una atención de calidad a los pacientes sobre los servicios de salud.

Su función principal se basa en el apoyo a las actividades operativas, tácticas y estratégicas dentro de una institución hospitalaria, por ejemplo, la atención al paciente eficiente y una alta satisfacción tanto de los pacientes como del personal (Winter et al., 2011).

Los sistemas de información hospitalarios se podrían clasificar en tres grupos (Ávila, 2016):

- Apoyo en la atención sobre prestaciones de servicios de salud (Front Office)

Son aplicativos que sirven de apoyo para la atención integral del paciente sobre las diversas prestaciones de servicios de salud que requieran, las mismas que demandan el contacto directo con el paciente.

- Apoyo al funcionamiento de la institución hospitalaria (Back Office)

Son aplicativos que gestionan las actividades de apoyo destinadas al funcionamiento de la institución hospitalaria, y que por su naturaleza no requieren contacto directo con el paciente.

- Sistemas de información clínicos

Son aplicativos que procesan información clínica del paciente de forma automática, empleando para realizar esta labor diversos tipos de integraciones entre sistemas hospitalarios.

2.10 SISTEMA DE INFORMACIÓN DE LABORATORIO CLÍNICO (LIS)

Conocido también como sistema de gestión de información de laboratorio, comprende la sección de la historia clínica electrónica, que se encarga de gestionar los datos del laboratorio que interopera con los datos clínicos del sistema de información hospitalario (Vargas-Herrera, 2015), en la actualidad se ha convertido en una herramienta indispensable para el desenvolvimiento de las actividades que realizan los laboratorios clínicos.

Laboratory Information System (LIS), sistemas de información de laboratorio por sus siglas en inglés, inicialmente fueron desarrollados para recolectar, registrar, presentar, organizar y archivar los resultados de laboratorio, con un enfoque en la generación de información para la gestión financiera de laboratorio; mientras las tecnologías de información, en general, están avanzando rápidamente, tanto en el dominio de hardware, como en el desarrollo de software, los LIS no han evolucionado correspondientemente; dado que, en la actualidad usan limitadamente la tecnología aplicada al diagnóstico, tales como mediciones electromédicas, algoritmos genéticos, isótopos radiactivos o técnicas de espectrometría de masa (De León, 2013), lo cual hace pensar que estos sistemas siguen una tendencia conservadora y resistente al cambio.

La complejidad de la información que producen los laboratorios clínicos se ha incrementado con el tiempo, y con el advenimiento de las técnicas empleadas por los analizadores a gran escala, como los microarrays y secuenciación de próxima generación, la cantidad de datos producidos crecerán rápidamente. Los desarrollos avanzados que se están aplicando en la gestión de datos y bioinformática deberán ser incorporados en el LIS para que este gran volumen de datos llegue a ser útil clínicamente (Sepulveda, Young, 2013). Con respecto al tratamiento asistencial, los LIS ayudan en los aspectos pre-analíticos (solicitud, cita, recolección de muestras, preparación, transporte, fraccionamiento y distribución), en el aspecto analítico (procesamiento de la muestra, gestión de equipos, control de calidad, validación técnica), así como en el aspecto pos-

analítico (validación clínica de la muestra, edición de informes, distribución, archivo de muestras).

En conjunto con el tratamiento asistencial, el LIS debe incorporar herramientas de gestión y comunicación (aprovisionamiento y gestión de almacenes, sistemas de calidad, páginas web). Además, debe integrarse en el conjunto de los sistemas de información de asistencia hospitalaria potenciando su utilidad (López, 2012).

2.11 LENGUAJES DE PROGRAMACIÓN WEB DE CÓDIGO ABIERTO

Entre los principales lenguajes de programación de código abierto exitosos tenemos tres de ellos que se destacan: Python, PHP y Java (Paredes, 2023).

Python, utilizado en una amplia gama de aplicaciones web.

PHP, esencial en el desarrollo de aplicaciones web.

Java, lenguaje predominante en el desarrollo de aplicaciones web.

A continuación, se exponen las principales características de estos lenguajes de programación de código abierto para el desarrollo de aplicaciones web.

2.11.1 PHP

PHP significa Preprocesador de hipertexto, y es un lenguaje del lado del servidor que puede generar HTML, procesar datos, comunicarse con bases de datos y realizar diversas operaciones lógicas (PHP, 2024).

PHP es ampliamente utilizado para el desarrollo de aplicaciones web, especialmente para sistemas de gestión de contenido, como WordPress, Drupal y Joomla.

PHP también se puede utilizar para el comercio electrónico, ya que puede manejar funcionalidades complejas, como carritos de compras, pasarelas de pago, gestión de inventario, procesamiento de pedidos y seguridad. PHP también es flexible, escalable y compatible con muchos servidores web y plataformas.

2.11.2 JAVA

Java es una plataforma informática de lenguaje de programación creada por Sun Microsystems en 1995. Ha evolucionado desde sus comienzos hasta impulsar una gran parte del mundo digital actual, ya que es una plataforma fiable en la que se crean muchos servicios y aplicaciones. Los nuevos e innovadores productos y servicios digitales diseñados para el futuro también siguen basándose en Java (Java, 2024).

Java es un lenguaje multiplataforma, orientado a objetos y centrado en la red que se puede utilizar como una plataforma en sí mismo.

Lenguaje de programación rápido, seguro y confiable para codificar cualquier tipo de aplicación, desde aplicaciones móviles y software empresarial hasta aplicaciones de macrodatos y tecnologías de servidores (Amazon Web Services, 2024).

2.11.3 PYTHON

Python es un lenguaje de programación ampliamente utilizado en aplicaciones web, desarrollo de software, ciencia de datos y machine learning (ML). Los desarrolladores usan Python porque es eficiente y fácil de aprender, además de que se puede ejecutar en muchas plataformas diferentes. El software Python se puede descargar gratis, se integra bien a todos los tipos de sistemas y aumenta la velocidad del desarrollo (Amazon Web Services, 2024).

Lenguaje de alto nivel de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código. Se trata de un lenguaje de programación multiparadigma, ya que soporta parcialmente la orientación a objetos, programación imperativa y, en menor medida, programación funcional.

Administrado por Python Software Foundation, posee una licencia de código abierto, denominada Python Software Foundation License. Python se clasifica constantemente como uno de los lenguajes de programación más populares (TIOBE, 2024).

En el siguiente apartado, se exponen teorías y conceptos referente a las herramientas de código abierto que predominan en el diseño de aplicaciones web.

2.12 HERRAMIENTAS PARA DISEÑO WEB DE CÓDIGO ABIERTO

Entre las principales herramientas para el diseño de aplicaciones web de código abierto se encuentra: NetBeans, Eclipse, WordPress, y Bootstrap.

2.12.1 NETBEANS

NetBeans IDE, es un entorno de desarrollo integrado libre, hecho principalmente para el lenguaje de programación Java. Existe además un número importante de módulos para extenderlo. Es un producto libre y gratuito sin restricciones de uso.

NetBeans IDE ofrece herramientas de primera clase para el desarrollo de aplicaciones web, corporativas, de escritorio y móviles con Java. Es el primer IDE en ofrecer soporte para las últimas versiones de JDK, Java EE y JavaFX. Proporciona descripciones generales inteligentes para ayudarle a comprender y gestionar sus aplicaciones, lo cual incluye soporte inmediato para tecnologías populares, como Maven.

NetBeans IDE contiene tecnologías innovadoras listas para usar y es el estándar en el desarrollo de aplicaciones, gracias a sus características integrales para el desarrollo de aplicaciones, las constantes mejoras en el editor de Java y el perfeccionamiento del rendimiento y la velocidad (NetBeans IDE, 2024).

2.12.2 ECLIPSE

Eclipse es una plataforma de software compuesto por un conjunto de herramientas de programación de código abierto multiplataforma para desarrollar el proyecto "Aplicaciones de Cliente Enriquecido", opuesto a las aplicaciones "Cliente-liviano" basadas en navegadores. Esta plataforma, típicamente ha sido usada para desarrollar entornos de desarrollo integrados (Eclipse IDE, 2024).

Eclipse fue desarrollado originalmente por IBM como el sucesor de su familia de herramientas para VisualAge. Eclipse es ahora desarrollado por la Fundación Eclipse. Fue liberado originalmente bajo la Common Public License, pero después fue re-licenciado bajo la Eclipse Public License. La Free Software Foundation ha dicho que ambas licencias son licencias de software libre, pero son incompatibles con Licencia pública general de GNU (GNU GPL).

2.12.3 WORDPRESS

WordPress es un sistema de gestión de contenidos (Content Management System o CMS) lanzado el 27 de mayo de 2003, enfocado a la creación de cualquier tipo de página web. Originalmente alcanzó una gran popularidad en la creación de blogs, para luego convertirse en una de las principales herramientas para la creación de páginas web comerciales.

WordPress está desarrollado en el lenguaje PHP para entornos que ejecuten MySQL y Apache, bajo licencia GPL y es software libre. Las causas de su enorme crecimiento son, entre otras, su licencia GPL, su facilidad de uso y sus características como gestor de contenidos. Aunque la mayor parte de la documentación existente se basa en la utilización del motor de bases de datos MySQL sobre el servidor web Apache, este sistema de gestión de contenidos puede también instalarse en entornos que ejecuten PostgreSQL, MariaDB y Nginx. Otro punto para considerar sobre su éxito y extensión es la enorme comunidad de desarrolladores y diseñadores, encargados de programarlo en su núcleo o creando complementos (llamados plugins) y plantillas (llamadas temas) para la comunidad (WordPress, 2024).

2.12.4 BOOTSTRAP

Bootstrap es una biblioteca multiplataforma o conjunto de herramientas de código abierto para diseño de sitios y aplicaciones web. Contiene plantillas de diseño con tipografía, formularios, botones, cuadros, menús de navegación y otros elementos de diseño basado en HTML y CSS, así como extensiones de JavaScript adicionales. Principalmente se basa en el desarrollo front-end.

En WordPress, puede instalarse como tema o usarse para el desarrollo de plugins o, incluso, dentro de ellos para estilizar sus funciones. El

propósito del framework es ofrecerle al usuario una experiencia más agradable cuando navega en un sitio web.

Bootstrap es el segundo proyecto más destacado en GitHub¹ y es usado por la NASA y la MSNBC, entre otras organizaciones.

En el apartado siguiente, se exponen teorías y conceptos con respecto a las bibliotecas de software de código abierto que predominan en el diseño y desarrollo de gráficos sobre aplicaciones web.

2.13 BIBLIOTECAS DE SOFTWARE GRÁFICAS PARA APLICACIONES WEB DE CÓDIGO ABIERTO

Entre las principales bibliotecas de software gráficas para el diseño y desarrollo de aplicaciones web de código abierto se encuentra: FusionCharts, Chart.js, y Highcharts.

2.13.1 FUSIONCHARTS

Es una biblioteca de gráficos JavaScript para aplicaciones web y empresariales, utilizada por más de 27 000 empresas y más de 750 000 desarrolladores en todo el mundo. Incluye más de 90 gráficos y más de 1000 mapas que transforman todos los datos comerciales en tableros de mando interactivos y significativos. FusionCharts, viene con complementos de código abierto para bibliotecas populares (jQuery), marcos (AngularJS y React) y lenguajes de programación (ASP.NET y

PHP). También posee tanto gráficos básicos (línea, columna, pastel, etc., 2D y 3D) como gráficos muy complejos (cascada, Gantt, candelabro, línea de zoom, etc.), se trata de la colección más exhaustiva de gráficos, widgets y mapas de JavaScript de la industria (FusionCharts, 2024).

FusionCharts, se usa en tableros de negocios para visualizar datos en diferentes dominios y funciones, incluidos Ventas, Finanzas, Marketing, Fabricación, Salud, Educación, Gobierno y Tecnologías de Información.

2.13.2 CHART.JS

Chart.js es una biblioteca JavaScript gratuita de código abierto para la visualización de datos, que admite 8 tipos de gráficos: barra, línea, área, circular, burbuja, radar, polar y dispersión.

Creado por el desarrollador web con sede en Londres Nick Downie en 2013, ahora es mantenido por la comunidad y es la segunda biblioteca de gráficos JS más popular en GitHub por la cantidad de estrellas después de D3.js, considerada significativamente más fácil de usar, aunque menos personalizable. Chart.js se representa en HTML5 Canvas y está ampliamente considerado como una de las mejores bibliotecas de visualización de datos (Chart.js, 2024).

2.13.3 HIGHCHARTS

Highcharts, es una biblioteca de gráficos JavaScript que ha sido diseñada para ayudar a los desarrolladores a agregar gráficos interactivos sobre cualquier plataforma. Con los wrappers, también puedes implementar Highcharts mediante los lenguajes de programación más populares (.Net, PHP, Python, R, Java), iOS y Android y marcos como Angular, Vue y React. Ofrece productos adicionales, como Highcharts Stock, Highcharts Maps y Highcharts Gantt. Los desarrolladores pueden crear gráficos de stock o líneas de tiempo generales, vincular datos a la geografía con mapas o visualizar proyectos en diagramas de Gantt (HighCharts, 2024).

Finalmente, sobre el apartado siguiente se exponen teorías y conceptos con respecto a las bibliotecas de software de código abierto que predominan en el desarrollo de aplicaciones web interactivas.

2.14 BIBLIOTECAS DE SOFTWARE INTERACTIVAS PARA APLICACIONES WEB DE CÓDIGO ABIERTO

Entre las principales bibliotecas interactivas para el desarrollo de aplicaciones web de código abierto se encuentran JavaScript y JQuery.

2.14.1 JAVASCRIPT

JavaScript (JS) es un lenguaje de programación interpretado, dialecto del estándar ECMAScript. Se define como orientado a objetos, basado en prototipos, imperativo, débilmente tipado y dinámico. Se usa principalmente del lado del cliente, implementado como parte de un

navegador web permitiendo mejoras en la interfaz de usuario y páginas web dinámicas y JavaScript del lado del servidor (Server-side JavaScript o SSJS).

Desde 2012, todos los navegadores modernos soportan completamente ECMAScript 5.1, una versión de JavaScript. Los navegadores más antiguos soportan por lo menos ECMAScript. Se diseñó con una sintaxis similar a C++ y Java,⁵⁶ aunque adopta nombres y convenciones del lenguaje de programación Java (JavaScript, 2024).

Tradicionalmente se venía utilizando en páginas web HTML para realizar operaciones y únicamente en el marco de la aplicación cliente, sin acceso a funciones del servidor. Actualmente es ampliamente utilizado para enviar y recibir información del servidor junto con ayuda de otras tecnologías como AJAX. JavaScript se interpreta en el agente de usuario al mismo tiempo que las sentencias van descargándose junto con el código HTML.

2.14.2 JQUERY

jQuery es una biblioteca multiplataforma de JavaScript, creada inicialmente por John Resig, que permite simplificar la manera de interactuar con los documentos HTML, manipular el árbol DOM, manejar eventos, desarrollar animaciones y agregar interacción con la técnica

AJAX a páginas web. Fue presentada el 14 de enero de 2006 en el BarCamp NYC.

jQuery, es software libre y de código abierto, posee un doble licenciamiento bajo la Licencia MIT y la Licencia Pública General de GNU v2, permitiendo su uso en proyectos libres y privados. Ofrece una serie de funcionalidades basadas en JavaScript que de otra manera requerirían de mucho más código, es decir, con las funciones propias de esta biblioteca se logran grandes resultados en menos tiempo y espacio.

La sintaxis de JQuery está diseñada para facilitar la navegación por un documento, seleccionar elementos DOM, crear animaciones, manejar eventos y desarrollar aplicaciones AJAX. JQuery también proporciona capacidades para que los desarrolladores creen complementos en la parte superior de la biblioteca de JavaScript. Esto permite a los desarrolladores crear abstracciones para interacción y animación de bajo nivel, efectos avanzados y widgets temáticos de alto nivel. El enfoque modular de la biblioteca jQuery permite la creación de poderosas páginas web dinámicas y aplicaciones web (jQuery, 2024).

A continuación, se expone la metodología que se aplicará para el desarrollo de la aplicación web integrando de manera interactiva los métodos de Biplot, mediante lenguajes de programación de código.

CAPÍTULO 3

3. METODOLOGÍA

3.1 INTRODUCCIÓN

Actualmente, existen aplicaciones de escritorio que permiten gestionar los métodos de la familia Biplot, sin embargo, es poco conocido sobre la existencia de aplicaciones web que cubran la misma finalidad. Por medio de la presente propuesta, desarrollaremos una Aplicación Web Interactiva que integre técnicas de Data Mining con los métodos de la familia Biplot: HJ-Biplot Clásico, Clustering HJ-Biplot, y Sparse Biplots, haciendo uso de herramientas de código abierto.

La metodología que permitirá cumplir con los objetivos planteados será la siguiente:

- Analizar, diseñar, y desarrollar la aplicación web interactiva.

Análisis de la Aplicación Web Interactiva

- Analizar los lenguajes de programación de código abierto web, para seleccionar el óptimo con respecto al desarrollo de la aplicación web interactiva.
- Analizar las herramientas de gestión de código abierto para la visualización de aplicaciones web, y seleccionar las que generen un mejor aporte al

desarrollo de la aplicación web interactiva. Entre las herramientas de gestión debemos identificar:

- Herramientas para el diseño de aplicaciones web
- Bibliotecas de software gráficas para aplicaciones web
- Bibliotecas de software interactivas para aplicaciones web
- Analizar las fuentes de datos para usarlas en la aplicación web interactiva.

Diseño de la Aplicación Web Interactiva

- Preparación del conjunto de datos sobre la etapa de preprocesamiento y la aplicación del proceso de ingeniería de variables.
- Diseñar las opciones y funcionalidades de la aplicación web interactiva.

Desarrollo de la Aplicación Web Interactiva

- Construir el proceso “Model Building”, para aplicar un modelo de clasificación o regresión al conjunto de datos, mediante los métodos de Biplot.
- Construir el proceso “Model Evaluation”, para seleccionar la técnica de validación de acuerdo con la medición y análisis del rendimiento de un modelo entrenado en un conjunto de datos.
- Codificar los programas que permitirá construir la aplicación web interactiva.

A continuación, en la Figura 3.1 se ilustra la metodología a implementar:

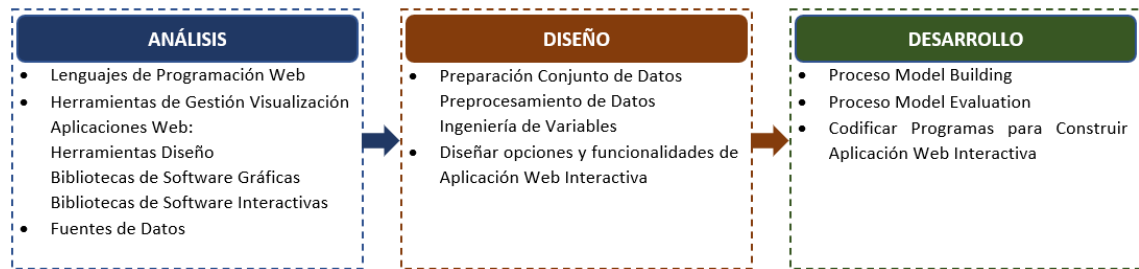


Figura 3.1. Metodología.

Fuente: Elaboración propia.

3.2 ANÁLISIS DE LA APLICACIÓN WEB INTERACTIVA

A continuación, se analizan elaborando un resumen de los lenguajes de programación de código abierto web, las herramientas de gestión de código abierto para la visualización de aplicaciones web y las fuentes de datos para usarlas en un entorno web interactivo.

3.2.1 LENGUAJE DE PROGRAMACIÓN WEB

En este apartado se analizan los lenguajes de programación de código abierto para el desarrollo de aplicaciones web, entre los cuales tenemos: PHP, Java, y Python.

PHP, lenguaje de programación cuyo propósito es establecer dinamismo a una web HTML estática. Puede usarse en diversas aplicaciones, sin embargo, su objetivo primordial es el desarrollo de aplicaciones web (PHP, 2024).

Java, uno de los lenguajes de programación más experimentados debido a que aparece en el año 1995, sigue siendo uno de los lenguajes de programación más populares. Funciona en cualquier sistema operativo, por consiguiente, se puede trasladar las aplicaciones a cualquier plataforma, permitiendo ejecutar las aplicaciones en sistemas más robustos a medida de sus necesidades (Java, 2024).

Python, permite escribir un programa con menos líneas de código en comparación con otros lenguajes. Está disponible una biblioteca estándar que contiene códigos reutilizables para casi cualquier tarea. Su comunidad activa incluye millones de desarrolladores alrededor del mundo. Puede trasladarse a través de diferentes sistemas operativos como Windows, macOS, Linux y Unix (Amazon Web Services, 2024).

PHP es el lenguaje de programación de código abierto que usaremos en el presente trabajo de titulación, debido a que su ventaja principal frente a los demás lenguajes de programación es que su objetivo principal es el desarrollo de aplicaciones web.

3.2.2 HERRAMIENTAS DE VISUALIZACIÓN DE APLICACIONES WEB

Se analizan las herramientas de gestión de código abierto para visualización de aplicaciones web, entre los cuales tenemos: herramientas para el diseño de aplicaciones web, bibliotecas de software

gráficas para aplicaciones web, y bibliotecas de software interactivas para aplicaciones web.

3.2.2.1 HERRAMIENTAS PARA DISEÑO DE APLICACIONES WEB

Se analiza y resume las principales herramientas de código abierto para el diseño de aplicaciones web.

NetBeans, posee amplia compatibilidad con diferentes lenguajes de programación, como Java, C++, PHP y HTML, lo cual lo convierte en una herramienta versátil para desarrollar distintos tipos de aplicaciones. Contiene una interfaz de usuario intuitiva y fácil de usar, lo que facilita la escritura de código y la navegación por el entorno de programación. NetBeans cuenta con una capacidad de depuración de código, que permite encontrar errores y corregirlos de manera eficiente. La principal desventaja es el consumo de recursos, ya que puede requerir una cantidad considerable de memoria RAM y procesador, que puede afectar el rendimiento del servidor de aplicaciones (NetBeans IDE, 2024).

Eclipse, entre las principales bondades encontramos la disponibilidad de un editor de texto con resaltador de sintaxis, posee una compilación de las aplicaciones en tiempo real, contiene un control de versiones (CVS) y un asistente (wizards)

para creación de proyectos. El entorno de desarrollo integrado (IDE) de Eclipse emplea módulos (plug-in) para proporcionar toda su funcionalidad al frente de la Plataforma de Cliente Enriquecido. Su principal inconveniente, es el alto consumo de recursos del sistema (Eclipse IDE, 2024).

Bootstrap, framework originalmente creado por Twitter, que permite crear interfaces web con CSS y JavaScript, cuya particularidad es la de adaptar la interfaz del sitio web al tamaño del dispositivo en que se visualice. Además, se basa en la simplicidad de sus interfaces, lo cual es una tendencia del mercado, en las que tiende a diseño plano, botones grandes, etc. para facilitar la usabilidad en los dispositivos más pequeño, atendiendo a la Experiencia de Usuario (Bootstrap, 2024).

WordPress, sistema de gestión de contenidos (CMS) muy popular de código abierto, ya que es fácil de usar y te permite hacer blogs o páginas web de todo tipo. Permite gestionar tu sitio web o blog de forma autónoma. Puedes personalizar la apariencia, diseño y contenido, incluyendo imágenes, videos y enlaces a redes sociales. Su manejo es intuitivo, y no es necesario ser un experto en la web para comenzar su uso. Entre las principales desventajas es que contiene vulnerabilidades de

seguridad y funcionalidades especializadas no están disponibles en un plugin o plantilla (WordPress, 2024).

Bootstrap, es la herramienta de diseño web de código abierto que usaremos en el presente trabajo de titulación, debido a que a diferencia de muchas herramientas de diseños web, su objetivo principal es el desarrollo de la interfaz de usuario (front-end), y posee una velocidad considerable en el desarrollo de aplicaciones web.

3.2.2.2 BIBLIOTECAS DE SOFTWARE INTERACTIVAS PARA APLICACIONES WEB

Las principales bibliotecas de código abierto para el desarrollo de aplicaciones web interactivas que usaremos son JavaScript y jQuery, cuyas características principales se las expone a continuación:

JavaScript ofrece interactividad en tiempo real, se refiere a la capacidad de una página web para responder y adaptarse a las acciones del usuario en el momento en que ocurren, sin la necesidad de recargar la página por completo. Esto crea una experiencia de usuario más dinámica, atractiva y fluida.

Mediante la técnica AJAX, esta herramienta carga y muestra contenido adicional en una página sin recurrir a una recarga.

JavaScript se utiliza para crear animaciones y efectos visuales en una página web. Gracias a esas funciones, puedes hacer que los elementos se muevan, cambien de tamaño, se desvanezcan y realicen otras funcionalidades de manera suave y atractiva (JavaScript, 2024).

jQuery ofrece una amplia variedad de animaciones y efectos visuales, con el fin de crear páginas web más atractivas e interactivas para los usuarios. Ayuda a añadir interacciones de usuario a las páginas web, como hacer clic en botones, mover el mouse, desplazarse por la página, etc. Asimismo, tiene una amplia gama de funciones para la validación de formularios, lo que ayuda a mejorar la experiencia del usuario al garantizar que los datos ingresados sean válidos antes de enviarlos al servidor (jQuery, 2024).

3.2.3 FUENTES DE DATOS

En este apartado se analizan las diversas fuentes de datos que usaremos sobre la aplicación web interactiva.

Entre las posibles fuentes de datos tenemos los archivos con formato Microsoft Excel, archivos con valores separados por comas (CSV), y archivo de texto sin formato. Los datos que estarán almacenados en los archivos se los puede extraer de las diversas fuentes de datos, entre las

que destacamos: base de datos, archivos planos, documentos con formato XML, datos de internet, datos sectoriales recopilados por empresas especializadas, etc.

Para el uso y demostración de la aplicación web interactiva tenemos información de la ciudad de Guayaquil, en el periodo comprendido entre 2019-2023 con una población accesible de todos los pacientes atendidos en la unidad de preadmisión de un hospital oncológico a través de la plataforma Intranet de un Sistema de Información Hospitalario (HIS); pacientes que para su atención cuentan con financiamiento propio o financiados tanto por la Red Pública Integral de Salud (IESS, ISSFA, ISSPOL y MSP), como por la Red complementaria. Esta información se encuentra almacenada en un gestor de base de datos Oracle, la cual extraeremos programando una consulta haciendo uso del Lenguaje de Consulta Estructurada (SQL) y finalmente cargaremos el resultado de la consulta SQL en un archivo con formato Microsoft Excel.

3.3 DISEÑO DE LA APLICACIÓN WEB INTERACTIVA

Prepararemos el conjunto de datos sobre la etapa de preprocesamiento y para la aplicación del proceso de ingeniería de variables, adicionalmente, diseñaremos las opciones y funcionalidades de la aplicación web interactiva.

3.3.1 PREPARACIÓN DEL CONJUNTO DE DATOS

Los datos provienen de pacientes atendidos desde la unidad de preadmisión de un Hospital Oncológico de Guayaquil - Ecuador, correspondientes a los años 2019-2023.

Aplicamos por medio de una consulta SQL el procesamiento de los datos para obtener el conjunto de los datos que usaremos en el presente trabajo. Hemos obtenido los lugares de nacimiento de los pacientes, su patología diagnóstica de cáncer, y las pruebas de laboratorio clínico. El conjunto de datos comprende 936 observaciones correspondientes a los lugares de nacimiento de los pacientes. A continuación, se ilustra el conjunto de los datos:

Pruebas de Laboratorio Clínico: son las variables que contienen los resultados de las pruebas de laboratorio clínico y están compuestos por 12 variables segregadas en tres grupos: Proteínas, Hematología y Bioquímica.

Proteínas

- TP: Proteínas Totales, medidos en g / dl (gramos por decilitro)
- IGS: Globulinas Sérica, medidos en g / dl (gramos por decilitro)
- ALBS: Albumina Sérica, medidos en g / dl (gramos por decilitro)

Hematología

- GB: Glóbulos Blancos, medidos en 10^3 / uL (microlitro)

- GR: Glóbulos Rojos, medidos en 10^6 / uL (microlitro)
- HGB: Hemoglobulina, medidos en pg (picogramo)
- HCT: Hematocrito, medidos en %
- PLT: Plaquetas, medidos en 10^3 / uL (microlitro)

Bioquímica

- CR: Creatinina, medidos en mg / dl (miligramos por decilitro)
- UREA: Urea, medidos en mg / dl (miligramos por decilitro)
- GOT: Glutamato Oxalacetato Transaminasa, medidos en U / l (unidades por litro)
- GPT: Glutamato Piruvato Transaminasa, medidos en U / l (unidades por litro)

Los individuos comprende tanto los lugares de nacimiento como la patología diagnóstica de cáncer de los pacientes. Adicionalmente, se dispone del año de procesamiento de las pruebas de laboratorio clínico.

Aplicando una ingeniería de variables, se ilustra en la Figura 3.2, el conjunto de datos seleccionado para el uso y demostración de la aplicación web interactiva.

AÑO	PROVINCIA	DIAGNOSTICO	TP	IGS	ALBS	GB	GR	HGB	HCT	PLT	CR	UREA	GOT	GPT
2019/01/01	AZUAY	C61 - TUMOR MALIGNO DE LA PRÓSTATA	7,55	2,95	4,6	6,62	3,6	11,85	35	163	1,61	71,75	24	11
2019/01/01	AZUAY	C50 - TUMOR MALIGNO DE LA MAMA	8,125	3,67	4,455	5,465	4,74	13,9	41,05	195,5	0,715	27,2	23	22
2019/01/01	AZUAY	C07 - TUMOR MALIGNO DE LA GLÁNDULA PARÓTI	7,4	2,64	4,76	4,78	4,12	13	39	298	1	28,3	20	19
2019/01/01	BOLIVAR	C61 - TUMOR MALIGNO DE LA PRÓSTATA	7,22	3,08	4,6	5,746666667	4,826666667	15,43333333	44,66666667	161,6666667	1,05	33,03333333	39	40
2019/01/01	BOLIVAR	C62 - TUMOR MALIGNO DEL TESTÍCULO	6,82	2,42	4,4	4,63	3,76	11,5	33	148	1,99	35,7	43	61
2019/01/01	BOLIVAR	C50 - TUMOR MALIGNO DE LA MAMA	7,723333333	2,92	4,803333333	6,9675	4,5725	13,525	40,95	219,25	0,7975	38,075	27	26,75
2019/01/01	BOLIVAR	C73 - TUMOR MALIGNO DE LA GLÁNDULA TIROID	7,12	2,81	4,31	6,79	4,75	13,9	40,4	291	0,52	19,6	12	11
2019/01/01	CAÑAR	C53 - TUMOR MALIGNO DEL CUELLO DEL ÚTERO	7,78	3,4	4,38	5,02	5,05	13,3	40,3	196	0,61	27,7	41	36
2019/01/01	CAÑAR	C61 - TUMOR MALIGNO DE LA PRÓSTATA	7,23	2,74	4,49	9,9625	3,775	12,575	36,85	180,5	2,9425	95,775	36,5	35
2019/01/01	CAÑAR	C833 - LINFOMA NO HODGKIN DE CÉLULAS GRAN	6,27	2,33	3,94	8,383333333	4,213333333	11,56666667	35,16666667	461,3333333	0,645	32,9	23,5	21
2019/01/01	CAÑAR	C82 - LINFOMA NO HODGKIN FOLICULAR (NODUL	7,33	2,47	4,86	7,14	4,54	13	39,5	290	0,74	23,9	23	34
2019/01/01	CAÑAR	C50 - TUMOR MALIGNO DE LA MAMA	8,06	3,32	4,74	8,27	4,436666667	13,86666667	39,56666667	287,6666667	0,72	22,725	24,5	24,5
2019/01/01	CAÑAR	C25 - TUMOR MALIGNO DEL PÁNCREAS	7,71	3,91	3,8	5,02	4,17	12,8	38,1	304	0,64	15	27	20
2019/01/01	CAÑAR	C67 - TUMOR MALIGNO DE LA VEJIGA URINARIA	7,39	3,02	4,36	11,84111111	3,205555556	9,088888889	26,34444444	441,8888889	2,04375	82,7625	24	35
2019/01/01	CARCHI	C73 - TUMOR MALIGNO DE LA GLÁNDULA TIROID	7,41	2,75	4,65	6,88	5,05	14,7	43	211	1,04	32,7	14	18
2019/01/01	CARCHI	C833 - LINFOMA NO HODGKIN DE CÉLULAS GRAN	7,11	2,47	4,64	4,39	4,04	12,7	38,9	226	0,72	39,4	21	10
2019/01/01	CARCHI	C184 - TUMOR MALIGNO DEL COLON TRANSVERS	7,71	3,04	4,67	5,85	4,78	14,4	42,4	210	0,78	29,2	19	15
2019/01/01	COTOPAXI	C50 - TUMOR MALIGNO DE LA MAMA	7,54	2,75	4,79	7,14	4,755	14,6	42,8	307,5	0,87	30,55	17	16
2019/01/01	COTOPAXI	C18 - TUMOR MALIGNO DEL COLON	7,77	3,28	4,49	4,99	5,66	16	47,8	222	1,02	44,25	47	54
2019/01/01	CHIMBORAZO	C838 - OTROS TIPOS ESPECIFICADOS DE LINFOMA	6,98	1,99	4,99	5,01	4,52	14	41,9	154	0,69	27,2	24	15
2019/01/01	CHIMBORAZO	C61 - TUMOR MALIGNO DE LA PRÓSTATA	7,21	2,765	4,445	7,493333333	3,776666667	10,8	32,26666667	286	1,03	57,03333333	23	16
2019/01/01	CHIMBORAZO	C16 - TUMOR MALIGNO DEL ESTOMAGO	5,2	2,59	2,61	18,955	3,37	9,3	28,45	154,5	1,68	112,2	400	191

Figura 3.2. Conjunto de Datos.

Fuente: Elaboración Propia.

3.3.2 DISEÑO DE OPCIONES Y FUNCIONALIDADES DE APLICACIÓN WEB INTERACTIVA

En este apartado se expone el diseño de las opciones y funcionales que tendrá disponible la aplicación web interactiva. Básicamente se dispone de tres secciones:

Archivo: contiene una sección que permite gestionar el archivo del conjunto de datos que se analizará y varias opciones que permiten configurar los gráficos que se emitirán por medio de la aplicación web interactiva.

Gráficas: contiene los gráficos que se generarán producto de las configuraciones aplicadas en la sección de Archivo.

Resumen: dispone de una sección que permite ilustrar un resumen de las funcionalidades ejecutadas por medio del lenguaje de programación R.

3.3.2.1 SELECCIÓN DE ARCHIVO

La aplicación web interactiva, permitirá usar archivos con formato Microsoft Excel, archivos con valores separados por comas (CSV), y archivo de texto sin formato, como se ilustra en la Figura 3.3.

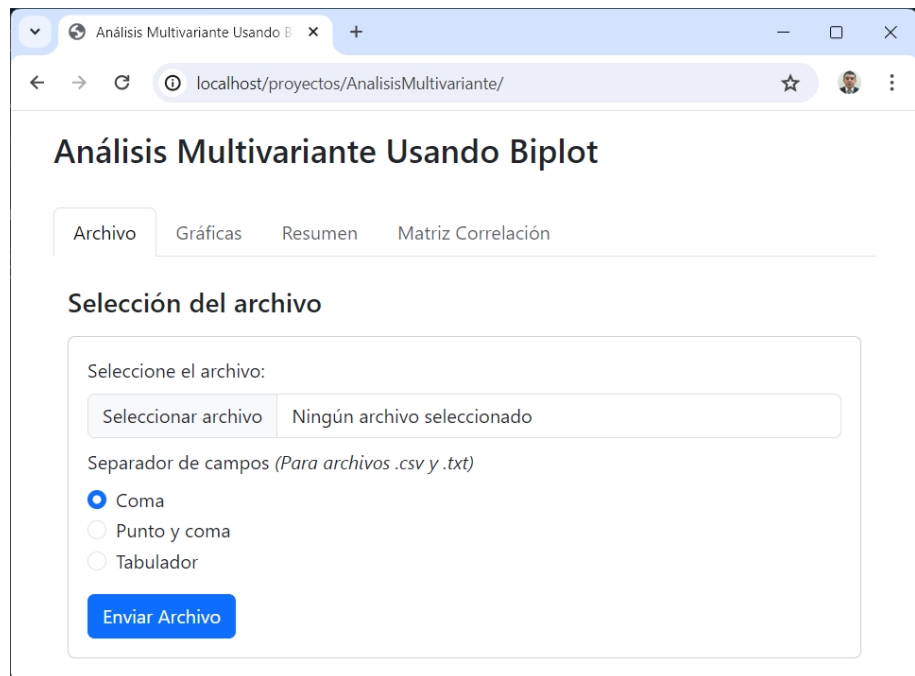


Figura 3.3. Selección de Archivo.

Fuente: Elaboración Propia.

3.3.2.2 ARCHIVO SELECCIONADO

Cuando se carga el archivo que usaremos para el análisis de los datos, como se ilustra en la Figura 3.4, la aplicación web interactiva permite:

- Ilustrar el contenido de los datos
- Identificar el tipo de variable de los datos (variable cuantitativa, variable cualitativa y variable de tipo fecha)
- Ilustrar la cantidad de filas y cantidad de columnas contenidas en el archivo

Archivo seleccionado

#	AÑO (Fecha)	PROVINCIA (Cual.)	DIAGNOSTICO (Cual.)	TP (Cuant.)	IGS (Cuant.)	ALBS (Cuant.)	GB (Cuant.)	GR (Cuant.)	HGB (Cuant.)
1	2019/01/01	AZUAY	C61 - TUMOR MALIGNO DE LA PRÓSTATA	7.55	2.95	4.6	6.62	3.6	11.85
2	2019/01/01	AZUAY	C50 - TUMOR MALIGNO DE LA MAMA	8.125	3.67	4.455	5.465	4.74	13.9
3	2019/01/01	AZUAY	C07 - TUMOR MALIGNO DE LA	7.4	2.64	4.76	4.78	4.12	13

Columnas: 15 Filas: 936

Figura 3.4. Archivo seleccionado.

Fuente: Elaboración Propia.

3.3.2.3 CONFIGURACIONES GENERALES

Para uso de todas las gráficas que se generan por medio de la aplicación web interactiva, tendremos que seleccionar las configuraciones generales, como se ilustra en la Figura 3.5:

- Estandarización: Permite aplicar escalamiento o estandarización sobre el conjunto de datos que se está analizando.
- Filas (Individuos / Observaciones): Se debe seleccionar la variable cualitativa sobre la cual se realizará el análisis de las variables cuantitativas que conforman el conjunto de datos.
- Columnas (Cuantitativas): Se debe seleccionar las variables cuantitativas que conforman el conjunto de datos, sobre las que se realizará el análisis multivariante.

Para todas la gráficas

Estandarización:

Si
 No

Filas (Individuos/Observaciones):

PROVINCIA ▼

Columnas (Cuantitativas):

TP
IGS
ALBS
GB
GR
HGB
HCT

Figura 3.5. Configuraciones Generales.

Fuente: Elaboración Propia.

3.3.2.4 CONFIGURACIONES PARA GRAFICAS DE LÍNEAS Y CIRCULAR

La aplicación web interactiva, como se aprecia en la Figura 3.6, permite configurar los parámetros necesarios para generar tanto gráficos de líneas como gráficos circulares.

- Gráficos de Líneas y Circular.

Individuo / Observación: de acuerdo con la fila (individuos / observaciones) seleccionada, se presentarán todos los individuos u observaciones que conforman ese conjunto de datos. Se debe seleccionar el individuo u observación sobre la cual se generará tanto la gráfica de líneas como la gráfica circular.

Título: Se permite imputar el título que se ilustrará el eje Y tanto para la gráfica de líneas como para la gráfica circular.

- Gráfica de Líneas: Se debe seleccionar la variable de tiempo que conforma el conjunto de datos de la variable X, sobre la cual se generará la gráfica de líneas.

The image shows two configuration panels. The top panel is titled 'Gráfica de Líneas y Circular' and contains a dropdown menu for 'PROVINCIA (Categoría)' with 'GUAYAS' selected, and a text input field for 'Título' containing 'Resultados Laboratorio Clínico'. The bottom panel is titled 'Gráfica de Líneas' and contains a dropdown menu for 'Variable de Tiempo' with 'AÑO' selected.

Figura 3.6. Configuraciones para Gráficos de Líneas y Circular.

Fuente: Elaboración Propia.

3.3.2.5 CONFIGURACIONES PARA MATRIZ DE CORRELACIÓN

La aplicación web interactiva, como se aprecia en la Figura 3.7, permite configurar los parámetros necesarios para generar la matriz de correlación de las variables.

The image shows a configuration panel titled 'Matriz Correlación:' with two radio button options: 'Si' (selected) and 'No'.

Figura 3.7. Configuraciones para Matriz de Correlación.

Fuente: Elaboración Propia.

3.3.2.6 CONFIGURACIONES PARA DIAGRAMA DE CAJAS

La aplicación web interactiva, permite configurar los parámetros necesarios para generar un diagrama de cajas, como se aprecia en la Figura 3.8.

- Diagrama de Cajas

Etiqueta Eje Y: se permite imputar el título que se ilustrará el eje Y del diagrama de cajas.

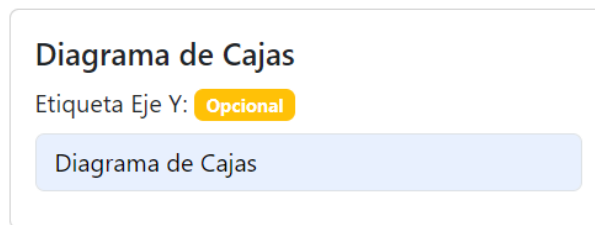


Diagrama de Cajas

Etiqueta Eje Y: **Opcional**

Diagrama de Cajas

Figura 3.8. Configuraciones para Diagrama de Cajas.

Fuente: Elaboración Propia.

3.3.2.7 CONFIGURACIONES PARA HJ BILOT

En este apartado se expone la configuración de los parámetros necesarios para la generación de un HJ Biplot, como se aprecia en la Figura 3.9.

- HJ Biplot

Color Puntos: se permite seleccionar el color con el que se graficarán los individuos u observaciones.

Color Etiquetas: se permite seleccionar el color con el que se graficarán las etiquetas de los individuos u observaciones.

Operación: se permite aplicar una suma o promedio de los individuos u observaciones que conforman el conjunto de datos analizados.

Escalamiento: se permite seleccionar el nivel de escalamiento para la transformación del conjunto de datos analizados. El objetivo de esta escalabilidad es balancear la representación de las observaciones y las variables en el gráfico.

Las posibles transformaciones del conjunto de datos son:

- 1.- "Raw Data": No se requiere transformación.
- 2.- "Subtract the global mean": Elimina un efecto común a todas las observaciones.
- 3.- "Double centering": Residuales de interacción. Cuando todos los elementos de la tabla sean comparables. Útil para modelos AMMI.
- 4.- "Column centering": Para quitar los medios de la columna.
- 5.- "Standardize columns": Elimina las medias de las columnas y divide por su desviación estándar.
- 6.- "Row centering": Para quitar los medios de fila.
- 7.- "Standardize rows": Divide cada fila por su desviación estándar.
- 8.- "Divide by the column means and center": La dispersión resultante es el coeficiente de variación.

9.- “Normalized residuals from independence” para una tabla de contingencia.

HJ Biplot y Cluster

Color Puntos:

Color Etiquetas:

Operación:

Escalamiento:

Figura 3.9. Configuraciones para HJ Biplot.

Fuente: Elaboración Propia.

El gráfico HJ Biplot, contiene las funcionalidades interactivas, como se aprecia en la Figura 3.10.

- **Mostrar etiquetas:** permite mostrar u ocultar las etiquetas tanto de los individuos/observaciones como de las variables.
- **Mostrar etiquetas:** permite mostrar u ocultar las etiquetas tanto de los individuos/observaciones como de las variables.
- **Sólo etiqueta cabecera:** permite mostrar u ocultar las etiquetas sólo de los individuos/observaciones. Chequear “Sólo etiqueta cabecera”, implica que se mostrará los valores contenidos en los individuos/observaciones.

- Representación 3D: permite ilustrar el gráfico HJ Biplot, en modelado 3D. Chequear “Representación 3D”, implica que se mostrará el gráfico en modelado 3D y des chequear implica que se mostrará el gráfico en modelado 2D.
- Eje X: permite ilustrar el gráfico HJ Biplot, de acuerdo con la selección de la dimensión que se considere para el Eje X.
- Eje Y: permite ilustrar el gráfico HJ Biplot, de acuerdo con la selección de la dimensión que se considere para el Eje Y.
- Eje Z: permite ilustrar el gráfico HJ Biplot, de acuerdo con la selección de la dimensión que se considere para el Eje Z.
- Contribución de Individuos/Observaciones: en un control deslizante (slider) que permite ilustrar el gráfico HJ Biplot, de acuerdo con las contribuciones de los individuos/observaciones sobre el eje que corresponde a la primera dimensión.
- Contribución de Variables: en un control deslizante (slider) que permite ilustrar el gráfico HJ Biplot, de acuerdo con las contribuciones de las variables sobre el eje que corresponde a la primera dimensión.

Mostrar etiquetas
 Sólo etiqueta cabecera
 Representación 3D

Eje X:
 Eje Y:
 Eje Z:

Contribución de Individuos/Observaciones entre: **0 - 88**

Contribución de Variables entre: **3 - 83**

Figura 3.10. Funcionalidades para HJ Biplot.

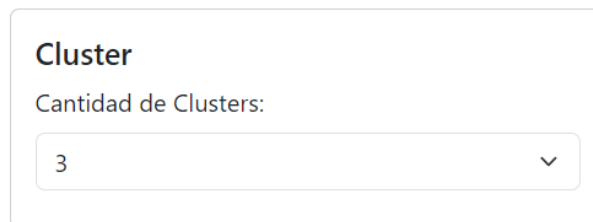
Fuente: Elaboración Propia.

3.3.2.8 CONFIGURACIONES PARA CLÚSTER EN HJ BILOT

En este apartado se expone la configuración de los parámetros necesarios para la generación de clusters en un HJ Biplot, como se aprecia en la Figura 3.11.

- Clúster en HJ Biplot

Cantidad de Clusters: se permite seleccionar la cantidad de clusters de los individuos u observaciones que conforman el conjunto de datos analizados para un HJ Biplot.



The image shows a configuration box titled "Cluster". Inside the box, there is a label "Cantidad de Clusters:" followed by a dropdown menu. The dropdown menu currently displays the number "3" and a downward-pointing arrow on the right side.

Figura 3.11. Configuraciones de Clusters para HJ Biplot.

Fuente: Elaboración Propia.

El gráfico de Clusters en HJ Biplot, contiene las funcionalidades interactivas expuestas en el apartado anterior 3.3.2.8, apreciados en la Figura 3.10.

3.3.2.9 CONFIGURACIONES PARA SPARSE BILOT

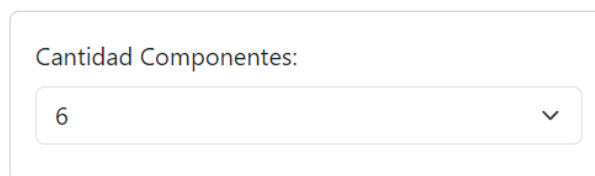
La aplicación web interactiva, permite configurar los parámetros necesarios para la generación de Sparse Biplot, como se aprecia en la Figura 3.12.

- Sparse Biplot

Cantidad de Componentes: se permite seleccionar la cantidad de componentes principales del Sparse Biplot para ilustrar la carga de las variables que conforman el conjunto de datos analizados.

Cardinalidad de Valores No Nulos: se permite seleccionar la cardinalidad de valores no nulos por cada componente principal del Sparse Biplot, para cada conjunto de datos analizado.

Penalización Lasso (λ): se permite seleccionar los valores para las penalizaciones Lasso (λ) por cada componente principal del Sparse Biplot, para cada conjunto de datos analizado.



Cantidad Componentes:

6

Cardinalidad de Valores No Nulas:

#1:	<input type="text" value="6"/>
#2:	<input type="text" value="4"/>
#3:	<input type="text" value="3"/>
#4:	<input type="text" value="2"/>
#5:	<input type="text" value="3"/>
#6:	<input type="text" value="2"/>

Penalización Lasso (lambda):

#1:	0.01	<input type="range" value="0.01"/>
#2:	0.16	<input type="range" value="0.16"/>
#3:	0.07	<input type="range" value="0.07"/>
#4:	0.02	<input type="range" value="0.02"/>
#5:	0.5	<input type="range" value="0.5"/>
#6:	0.5	<input type="range" value="0.5"/>

Figura 3.12. Configuraciones para Sparse Biplot.

Fuente: Elaboración Propia.

3.4 DESARROLLO DE LA APLICACIÓN WEB INTERACTIVA

En este apartado se expone el desarrollo de la aplicación web interactiva sobre la que se construirá el proceso “Model Building”, para aplicar un modelo de

clasificación o regresión al conjunto de datos analizados, mediante los métodos de Biplot. También, se construirá el proceso “Model Evaluation”, para seleccionar la técnica de validación de acuerdo con la medición y análisis del rendimiento de un modelo entrenado en el conjunto de datos analizado. Finalmente, partiendo de la construcción de los modelos se codificarán los programas que permitirá la construcción de la aplicación web interactiva.

3.4.1 INTERFAZ WEB DE VISUALIZACIÓN

Se codifica una interfaz web de visualización tipo Dashboard, usando el framework Bootstrap y el lenguaje de programación de código abierto PHP, como se ilustra en la Figura A1 del Anexo.

3.4.2 GESTIÓN DE ARCHIVOS

Se elabora la codificación en el lenguaje de programación de código abierto PHP, para gestionar archivos con formato Microsoft Excel, archivos con valores separados por comas (CSV), y archivo de texto sin formato, que permita analizar el conjunto de datos, como se aprecia en la Figura A2 del Anexo.

3.4.3 GRÁFICOS INTERACTIVOS Y DINÁMICOS

Por medio del conjunto de datos seleccionado en la Gestión de Archivos, se codifica los programas en el lenguaje de programación de código abierto PHP para la visualización de gráficos interactivos y dinámicos de

líneas, circulares y diagramas de cajas haciendo uso de la biblioteca gráfica JavaScript de código abierto denominada Highcharts, ilustrados en la Figura A3 del Anexo y sobre la Figura 3.13.

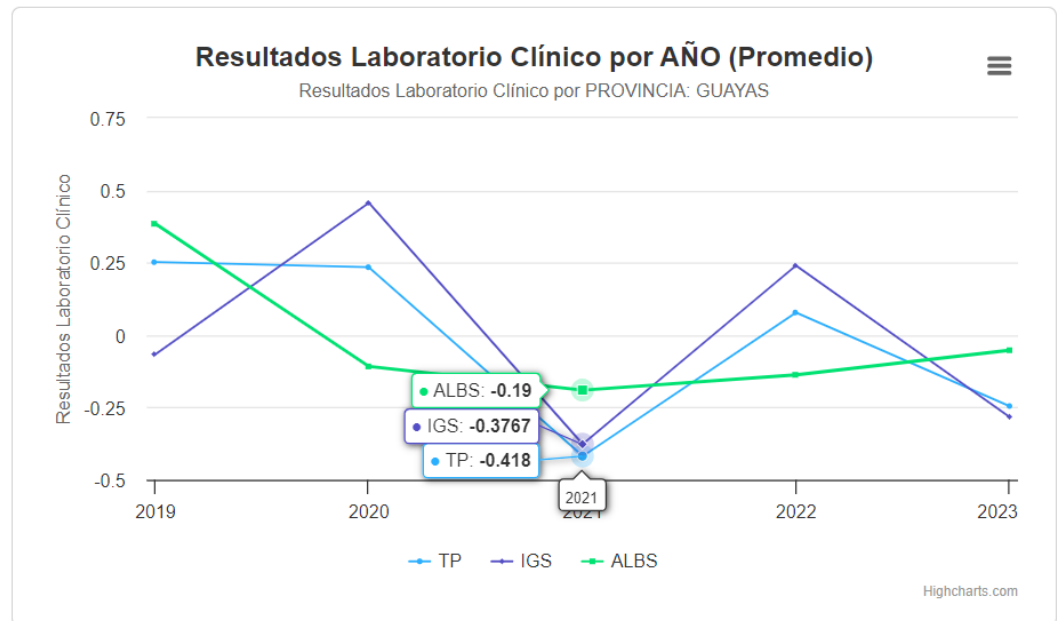


Figura 3.13. Gráfico de Línea Interactivo y Dinámico.

Fuente: Elaboración Propia.

3.4.4 GRÁFICOS PARA VISUALIZACIÓN EXPLORATORIA

Por medio del conjunto de datos seleccionado en la Gestión de Archivos, se codifican los programas en el lenguaje de programación de código abierto PHP para la visualización de gráficos interactivos y dinámicos exploratorios haciendo uso de la biblioteca gráfica JavaScript de código abierto denominada Highcharts, ilustrados en la Figura 3.14.

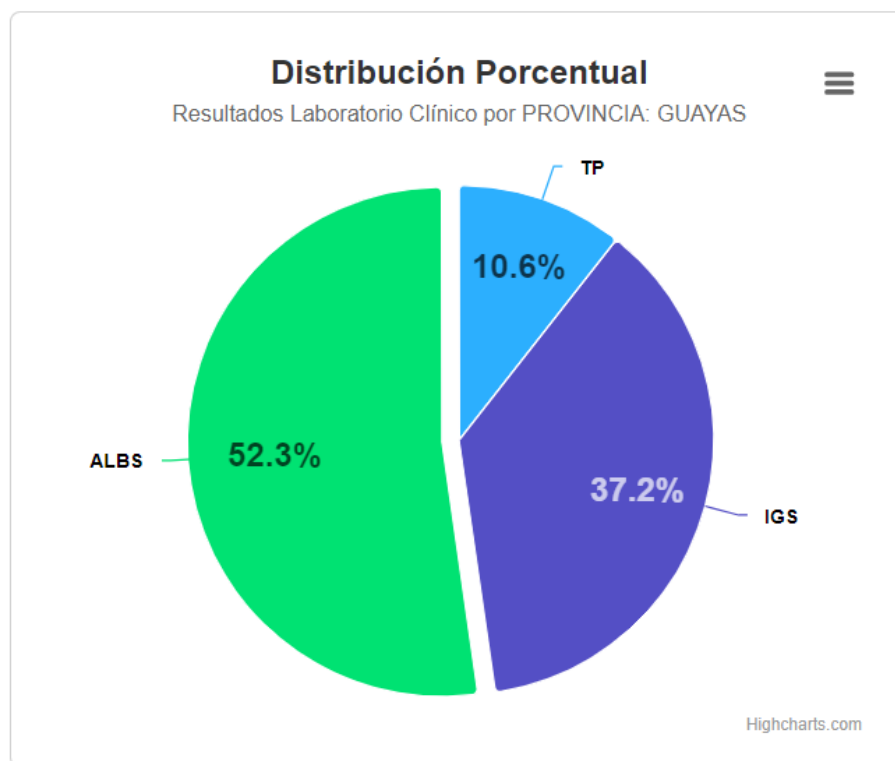


Figura 3.14. Gráfico Circular Exploratorio Interactivo y Dinámico.

Fuente: Elaboración Propia.

Se integran los programas por medio de la compilación del código usando el lenguaje de programación PHP. El programa codificado en PHP invoca a los programas codificados en el lenguaje de programación R. Finalmente, los programas codificados en R retornan como resultado de la invocación texto en formato de intercambio de datos JSON. Esta integración de lenguajes de programación se aprecia en la Figura A4 del Anexo.

3.4.5 GRÁFICOS PARA VISUALIZACIÓN AVANZADA

A través del conjunto de datos seleccionado en la Gestión de Archivos, se codifica los programas en el lenguaje de programación de código abierto PHP para la visualización de gráficos interactivos y dinámicos avanzados haciendo uso de la biblioteca gráfica JavaScript de código abierto denominada Highcharts, ilustrados en la Figuras 3.15. Entre gráficos interactivos y dinámicos avanzados se encuentran HJ Biplot, y Sparse Biplot.

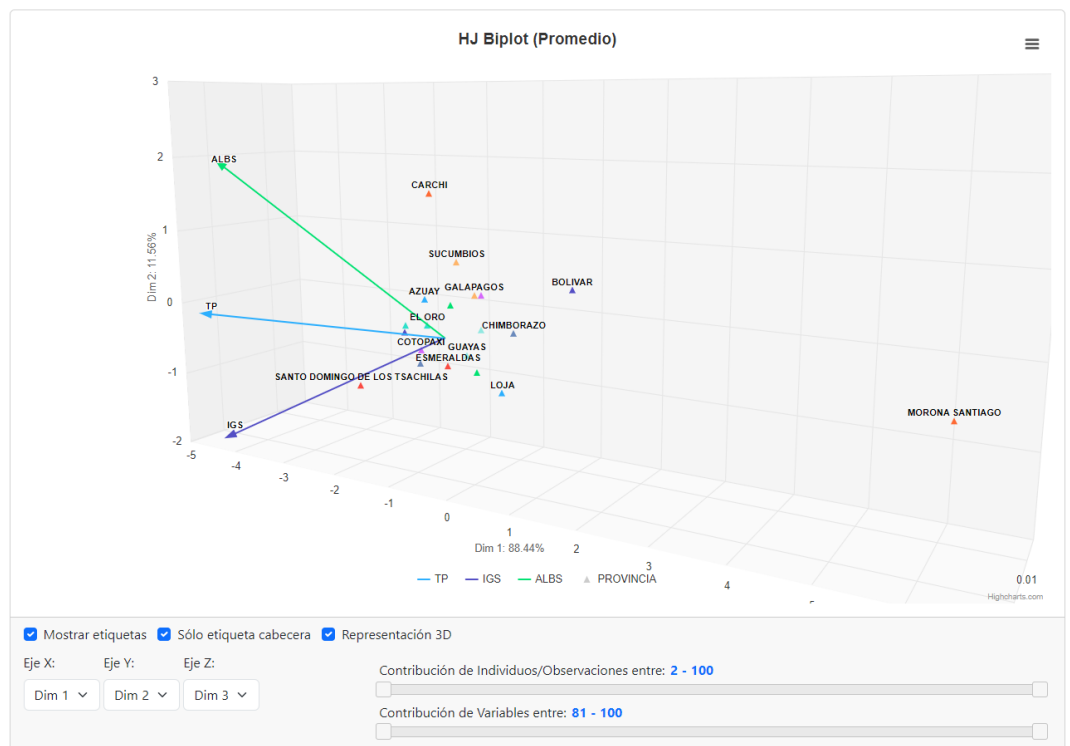


Figura 3.15. HJ Biplot Interactivo y Dinámico.

Fuente: Elaboración Propia.

Se integran los programas por medio de la compilación del código usando el lenguaje de programación PHP. El programa codificado en PHP invoca

a los programas codificados en el lenguaje de programación R. Finalmente, los programas codificados en R retornan como resultado de la invocación texto en formato de intercambio de datos JSON.

Sobre la integración expuesta se codifica los programas construyendo el proceso “Model Building”, aplicando un modelo de clasificación al conjunto de datos analizados, mediante el método HJ Biplot. También, se construye el proceso “Model Evaluation”, seleccionando el algoritmo de agrupamiento clustering de acuerdo con las mediciones y análisis del rendimiento del modelo entrenado en el conjunto de datos analizado. Adicionalmente, partiendo de la construcción de los modelos, se codificarán los programas en los lenguajes de programación PHP y R, que permitirá aplicar un Análisis de Componentes Principales Sparse SPCA, ilustrado en la Figura A5 del Anexo.

CAPÍTULO 4

4. RESULTADOS

4.1 INTRODUCCIÓN

La elaboración de este capítulo, tiene como finalidad exponer el análisis de los resultados que se generan producto de la emisión de resúmenes de información y gráficos exploratorios y avanzados por medio de la aplicación web interactiva, teniendo como fuente de información un archivo en formato Microsoft Excel que contiene un conjunto de datos de las pruebas de laboratorio clínico, en un periodo comprendido entre 2019 – 2023; sobre pacientes con patología diagnóstica de cáncer de un Hospital Oncológico de Guayaquil - Ecuador.

4.2 VISUALIZACIÓN EXPLORATORIA INTERACTIVA

En este apartado se presentan los resultados sobre los gráficos interactivos y dinámicos de líneas, circulares y diagramas de cajas.

4.2.1 GRAFICOS DE LÍNEAS

Se ilustra en la Figura 4.1 el gráfico de líneas interactivo y dinámico. Se observa que el resultado de las pruebas de laboratorio clínico de Glóbulos Rojos (GR), en pacientes con patología diagnóstica de cáncer

que nacieron en la provincia del Guayas en el periodo 2023, tiene un promedio de 4.75 millones de células por microlitro.

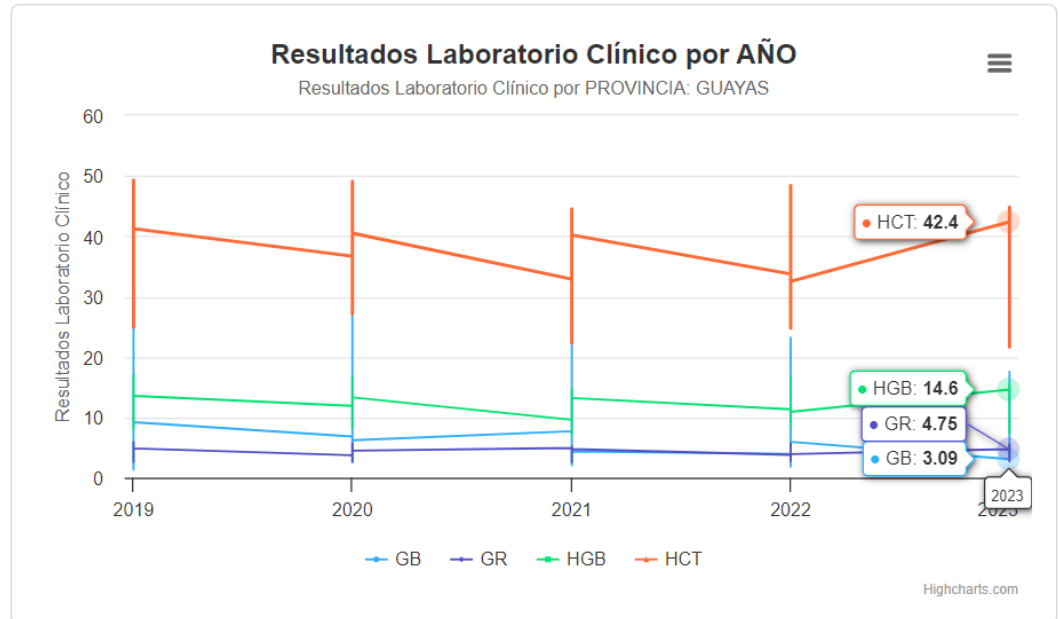


Figura 4.1. Gráfico de Línea Interactivo y Dinámico.

Fuente: Elaboración Propia.

4.2.2 GRAFICOS CIRCULARES

En la Figura 4.2 se presenta el gráfico circular interactivo y dinámico. Se observa que los resultados de las pruebas de laboratorio clínico de Hematocritos (HCT), en pacientes con patología diagnóstica de cáncer que nacieron en la provincia de Manabí en el periodo entre 2019-2023, representan el 10.8 % del grupo de pruebas de Hematología.

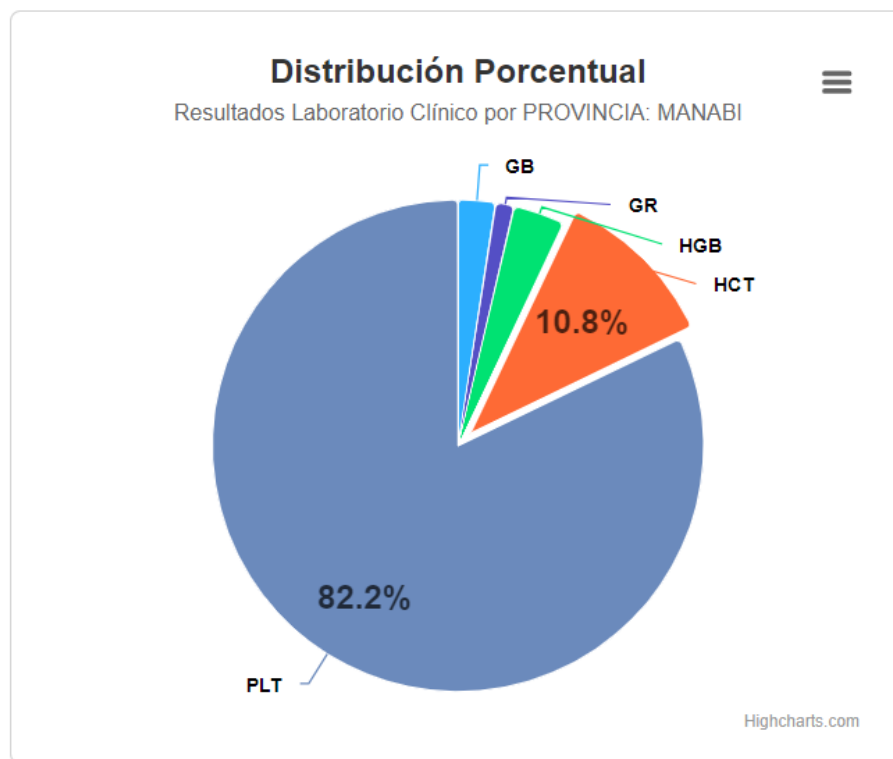


Figura 4.2. Gráfico Circular Interactivo y Dinámico.

Fuente: Elaboración Propia.

4.2.3 DIAGRAMA DE CAJAS

Se aprecia en el diagrama de cajas interactivo y dinámico de la Figura 4.3, que los resultados de las pruebas de laboratorio clínico de Hemoglobulina (HGB), en pacientes con patología diagnóstica de cáncer que nacieron en la provincia El Oro en el periodo entre 2019-2023, tienen en promedio un valor mínimo de 8.3 picogramos, valor de la mediana 12.07 picogramos y como valor máximo 14.95 picogramos.

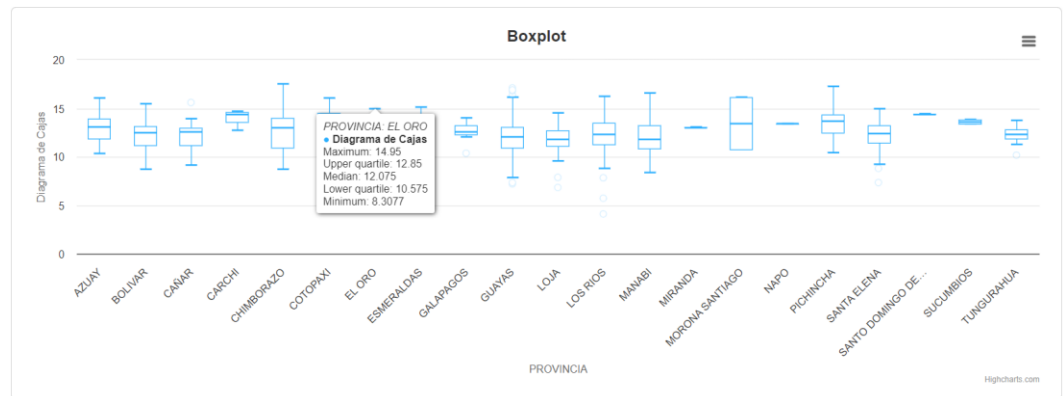


Figura 4.3. Diagrama de Cajas Interactivo y Dinámico.

Fuente: Elaboración Propia.

4.3 VISUALIZACIÓN AVANZADA INTERACTIVA

En este apartado se presentan los resultados sobre los gráficos interactivos y dinámicos de HJ Biplot, y Análisis de Componentes Principales Sparse.

4.3.1 HJ BILOT

En la Figura 4.4 del HJ Biplot interactivo y dinámico, se observa:

- Valores Propios

El primer eje con un valor de 101.27 (Dim 1).

El segundo eje con un valor de 54.39 (Dim 2).

El tercer eje con un valor de 36.85 (Dim 3).

- Variabilidad

De la inercia o variabilidad que se proyectará por medio de un gráfico, el primer eje logrará proyectar el 42.19% de la variabilidad, mientras que el segundo eje logrará proyectar el 22.66% de la variabilidad y el

tercer eje logrará proyectar el 15.35% de la variabilidad que representa la relación o estructura en común que existe entre los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.

Entre el primer y segundo eje se logrará proyectar el 64.86% de la estructura en común que existe entre los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.

- Las variables Hemoglobulina (HGB) y Hematocrito (HCT) tienen una fuerte correlación directa positiva, se entiende que si aumentan los valores de Hemoglobulina (HGB) también aumentarán los valores sobre las pruebas de laboratorio Hematocrito (HCT).
- La variable Creatinina (CR) tienen una fuerte correlación inversa con Glutamato Oxalacetato Transaminasa (GOT), se entiende que si aumentan los valores de Creatinina (CR) disminuirán los valores sobre las pruebas de laboratorio Glutamato Oxalacetato Transaminasa (GOT).
- Las variables Glutamato Oxalacetato Transaminasa (GOT) y Glutamato Piruvato Transaminasa (GPT), tienen una fuerte correlación directa positiva, se entiende que si aumentan los valores de Glutamato Oxalacetato Transaminasa (GOT) también aumentarán los valores sobre las pruebas de laboratorio Glutamato Piruvato Transaminasa (GPT).

- Las variables Globulinas Sérica (IGS), Albumina Sérica (ALBS) y Proteínas Totales (TP) tienen una fuerte correlación inversa con Glóbulos Blancos (GB). Se entiende que si aumentan los valores de Globulinas Sérica (IGS), Albumina Sérica (ALBS) y Proteínas Totales (TP), consecuentemente disminuirán los valores sobre las pruebas de laboratorio Glóbulos Blancos (GB).
- La variable Globulinas Sérica (IGS) no tiene injerencia sobre las pruebas de laboratorio Hemoglobulina (HGB).
- Los pacientes que nacieron en las provincias Los Ríos y Manabí tienen un comportamiento similar en relación con los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.
- Los pacientes que nacieron en la provincia Morona Santiago tienen un comportamiento diferente al resto de provincias en relación con los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.
- Los pacientes que nacieron en las provincias El Oro, Esmeraldas y Santa Elena tienen un comportamiento similar en relación con los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.
- Los pacientes que nacieron en la provincia Loja tienen un comportamiento diferente al resto de provincias en relación con los

resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.

- En los pacientes que nacieron en las provincias Loja y Guayas la estructura es baja entre los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, con patología diagnóstica de cáncer.
- La provincia de nacieron de pacientes Pichincha tiene una preferencia en relación con los resultados de las pruebas de Hemoglobulina (HGB).
- Los pacientes que nacieron en la provincia Sucumbíos tiene una preferencia en relación con los resultados de las pruebas de laboratorio Creatinina (CR).
- Los pacientes que nacieron en las provincias Azuay y Tungurahua tienen una preferencia en relación con los resultados de las pruebas de laboratorio Globulinas Sérica (IGS).

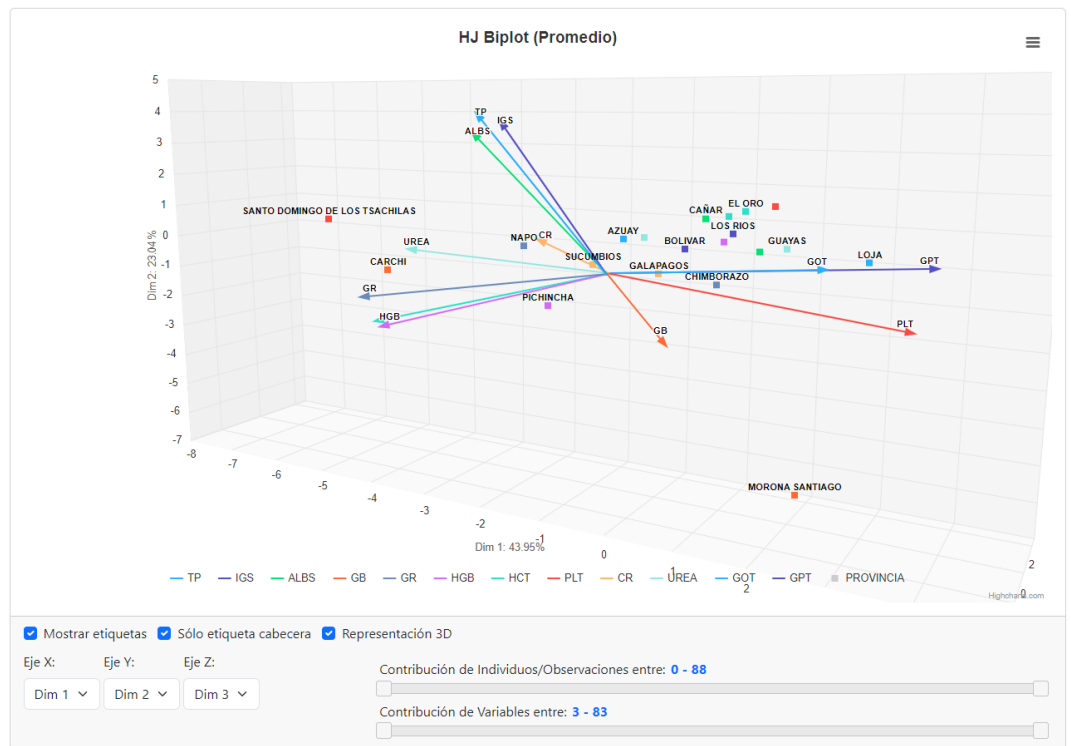


Figura 4.4. HJ Biplot Interactivo y Dinámico en 3D.

Fuente: Elaboración Propia.

En la Figura 4.5 se proyecta el HJ Biplot interactivo y dinámico en 2D, generado producto de analizar el conjunto de datos en pacientes de sexo femenino, apreciando una mejor representación de las variables y observaciones en el gráfico HJ Biplot, sobre el cual se observa:

- Variabilidad

De la inercia o variabilidad que se proyectará por medio de un gráfico, el primer eje logrará proyectar el 36.07% de la variabilidad, mientras que el segundo eje logrará proyectar el 17.9% de la variabilidad y el tercer eje logrará proyectar el 13.77% de la variabilidad que

representa la relación o estructura en común que existe entre los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes de sexo femenino con patología diagnóstica de cáncer.

Entre el primer y segundo eje se logrará proyectar el 53.97% de la estructura en común que existe entre los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes de sexo femenino con patología diagnóstica de cáncer.

- Las variables Urea (UREA), Creatinina (CR) y Glóbulos Blancos (GB) tienen una fuerte correlación directa positiva, se entiende que si aumentan los valores de Urea (UREA) también aumentarán los valores sobre las pruebas de laboratorio de Creatinina (CR) y Glóbulos Blancos (GB).
- Las variables Hemoglobulina (HGB) y Hematocrito (HCT) tienen una fuerte correlación directa positiva, se entiende que si aumentan los valores de Hemoglobulina (HGB) también aumentarán los valores sobre las pruebas de laboratorio Hematocrito (HCT).
- La variable Creatinina (CR) tienen una fuerte correlación inversa con Plaquetas (PLT), se entiende que si aumentan los valores de Creatinina (CR) disminuirán los valores sobre las pruebas de laboratorio Plaquetas (PLT).
- Las variables Glutamato Oxalacetato Transaminasa (GOT) y Glutamato Piruvato Transaminasa (GPT), tienen una fuerte

correlación directa positiva, se entiende que si aumentan los valores de Glutamato Oxalacetato Transaminasa (GOT) también aumentarán los valores sobre las pruebas de laboratorio Glutamato Piruvato Transaminasa (GPT).

- Las variables Glóbulos Rojos (GR), Hemoglobulina (HGB), Hematocrito (HCT) y Proteínas Totales (TP), tienen una fuerte correlación directa positiva, se entiende que si aumentan los valores de Glóbulos Rojos (GR) también aumentarán los valores sobre las pruebas de laboratorio Hemoglobulina (HGB), Hematocrito (HCT) y Proteínas Totales (TP).
- La variable Globulinas Sérica (IGS) no tiene injerencia sobre las pruebas de laboratorio Glóbulos Rojos (GR).
- Los pacientes de sexo femenino con diagnóstico patológico C90 - MIELOMA MÚLTIPLE Y TUMORES MALIGNOS DE CÉLULAS PLASMÁTICAS, y C101 - TUMOR MALIGNO DE LA CARA ANTERIOR DE LA EPIGLOTIS, tienen un comportamiento similar en relación con los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023.
- Los pacientes de sexo femenino con diagnóstico patológico C90 - MIELOMA MÚLTIPLE Y TUMORES MALIGNOS DE CÉLULAS PLASMÁTICAS, y C101 - TUMOR MALIGNO DE LA CARA ANTERIOR DE LA EPIGLOTIS, tienen un comportamiento similar en

relación con los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023.

- Los pacientes de sexo femenino con diagnóstico patológico D27 - TUMOR BENIGNO DEL OVARIO, tienen un comportamiento diferente al resto de diagnósticos patológicos en relación con los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023.
- Los pacientes de sexo femenino con diagnóstico patológico C540 - TUMOR MALIGNO DEL ISTMO UTERINO y C161 - TUMOR MALIGNO DEL FUNDUS GÁSTRICO, la estructura es baja entre los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023.
- Los pacientes de sexo femenino con diagnóstico patológico C227 - OTROS CARCINOMAS ESPECIFICADOS DEL HÍGADO, tiene una preferencia en relación con los resultados de las pruebas de Urea (UREA).
- Los pacientes de sexo femenino con diagnóstico patológico C444 - TUMOR MALIGNO DE LA PIEL DEL CUERO CABELLUDO Y DEL CUELLO, tiene una preferencia en relación con los resultados de las pruebas de Glóbulos Rojos (GR), Hemoglobulina (HGB), Hematocrito (HCT) y Proteínas Totales (TP).
- Los pacientes de sexo femenino con diagnóstico patológico C40 - TUMOR MALIGNO DE LOS HUESOS Y DE LOS CARTÍLAGOS

ARTICULARES DE LOS MIEMBROS, tiene una preferencia en relación con los resultados de las pruebas de Albumina Sérica (ALBS).

- Los pacientes de sexo femenino con diagnóstico patológico C779 - TUMOR MALIGNO DEL GANGLIO LINFÁTICO, SITIO NO ESPECIFICADO, C229 - TUMOR MALIGNO DEL HÍGADO, NO ESPECIFICADO y C22 - TUMOR MALIGNO DEL HÍGADO Y DE LAS VÍAS BILIARES INTRAHEPÁTICAS, tiene una preferencia en relación con los resultados de las pruebas de Glutamato Piruvato Transaminasa (GPT).

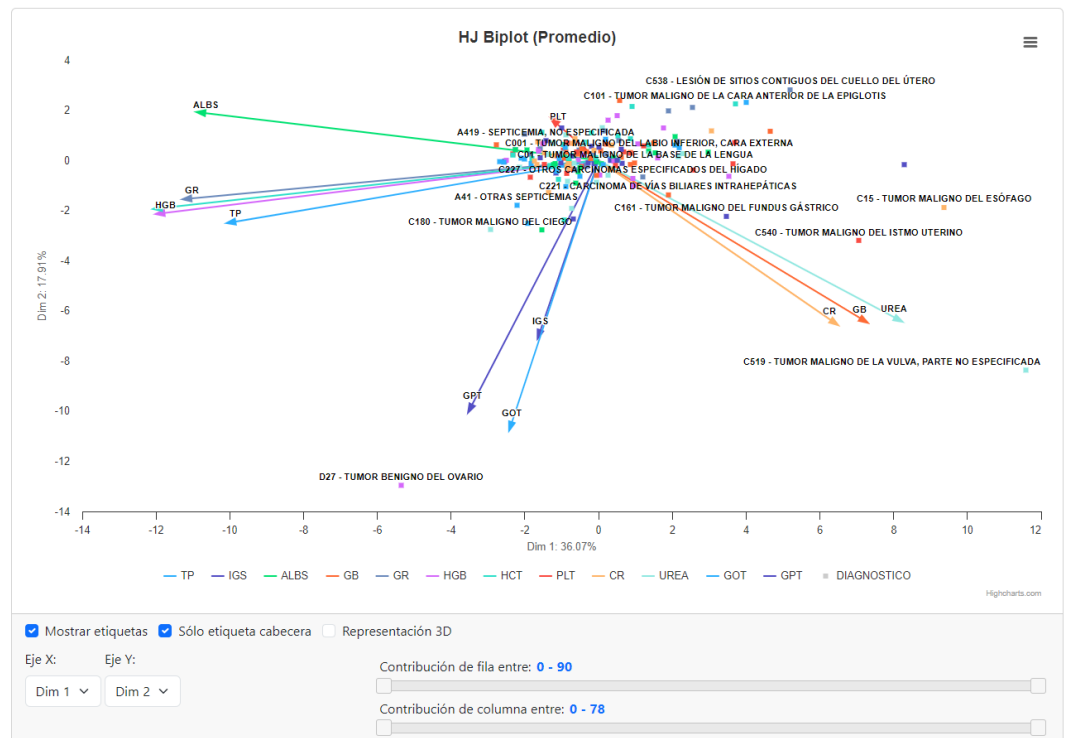


Figura 4.5. HJ Biplot Interactivo y Dinámico en 2D.

Fuente: Elaboración Propia.

Conforme a la ilustración de la Figura 4.6, se aprecia la contribución de las variables sobre las dimensiones del HJ Biplot.

Se puede interpretar que las variables que conformen el perfil Hematológico “Glóbulos Blancos (GB), Glóbulos Rojos (GR), Hemoglobulina (HGB), y Hematocrito (HCT)”, son las que se representan mejor en la dimensión 1, mientras que Glutamato Oxalacetato Transaminasa (GOT) y Glutamato Piruvato Transaminasa (GPT) se representan mejor en la dimensión 2; así mismo, se puede interpretar que la variables Urea (UREA) y Creatinina (CR) se representan mejor en la dimensión 3.

Column Contributions			
	Dim 1	Dim 2	Dim 3
TP	52.28	6.39	0.02
IGS	0.14	26.85	0.71
ALBS	75.85	0.95	0.65
GB	32.53	0.08	9.58
GR	69.87	2.50	8.35
HGB	77.36	2.96	11.86
HCT	78.72	3.06	10.98
PLT	0.19	6.34	30.93
CR	19.06	3.82	48.74
UREA	25.38	2.77	55.68
GOT	0.42	73.64	9.26
GPT	3.89	68.20	8.83

Figura 4.6. Contribución de Variables sobre las Dimensiones.

Fuente: Elaboración Propia.

4.3.2 CLÚSTER EN HJ BILOT

En la Figura 4.7 del Clúster en HJ Biplot interactivo y dinámico, se exponen las siguientes observaciones:

- Importancia de las componentes

La importancia de las componentes principales radica en su capacidad para explicar la varianza en los datos y para ayudar a identificar la relación entre las variables y las observaciones.

La desviación estándar de la primera componente principal representa el 1.24 de desviación de sus valores de una variable sobre su medida.

La desviación estándar de la segunda componente principal representa el 0.96 de desviación de sus valores de una variable sobre su medida.

La desviación estándar de la tercera componente principal representa el 0.6 de desviación de sus valores de una variable sobre su medida.

La desviación estándar de la cuarta componente principal representa el 0.34 de desviación de sus valores de una variable sobre su medida.

- Variabilidad

De la inercia o variabilidad que se proyectará por medio de un gráfico de clúster en HJ Biplot, el primer eje logrará proyectar el 49.79% de la variabilidad, mientras que el segundo eje logrará proyectar el 30.19% de la variabilidad y el tercer eje logrará proyectar el 11.91% de la variabilidad que representa la relación o estructura en común que existe entre los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.

Entre el primer y segundo eje se logrará proyectar el 79.99% de la estructura en común que existe entre los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.

- Las variables Globulinas Sérica (IGS), Albumina Sérica (ALBS) y Proteínas Totales (TP) tienen una fuerte correlación directa positiva. Se entiende que si aumentan los valores de Globulinas Sérica (IGS), también aumentarán los valores de laboratorio clínico Albumina Sérica (ALBS) y Proteínas Totales (TP).
- La variable Glóbulos Rojos (GR) tiene una fuerte correlación inversa con Glutamato Piruvato Transaminasa (GPT), se entiende que si aumentan los valores de Glóbulos Rojos (GR) disminuirán los valores de laboratorio clínico Glutamato Piruvato Transaminasa (GPT).
- La variable Glóbulos Blancos (GB) no tiene injerencia con las pruebas de laboratorio Plaquetas (PLT).

- Las variables Glutamato Piruvato Transaminasa (GPT), Glóbulos Rojos (GR) y Plaquetas (PLT) contienen las desviaciones estándar más altas, lo que implica que tienen una mayor contribución en la estructura de los datos.
- Los pacientes que nacieron en la provincia Santo Domingo De Los Tsáchilas conforman el primer conglomerado (clúster). Tienen un comportamiento diferente al resto de provincias en relación con los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.
- Los pacientes que nacieron en la provincia Morona Santiago conforman el segundo conglomerado (clúster). Los resultados de laboratorio clínico de estos pacientes tienen un comportamiento diferente al resto de provincias en relación con los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.
- Los pacientes que nacieron en las provincias Los Ríos, Manabí, Azuay, Bolívar, y Loja tienen un comportamiento similar en relación con los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.
- Los pacientes que nacieron en las provincias El Oro, Esmeraldas, Santa Elena, y Guayas tienen un comportamiento similar en relación con los resultados de las pruebas de laboratorio clínico sobre el

periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.

- Los pacientes que nacieron en la provincia Loja tienen una preferencia en relación con los resultados de las pruebas de laboratorio clínico Glutamato Piruvato Transaminasa (GPT).
- Los pacientes que nacieron en la provincia Loja tienen una preferencia en relación con los resultados de las pruebas de laboratorio clínico Creatinina (CR).

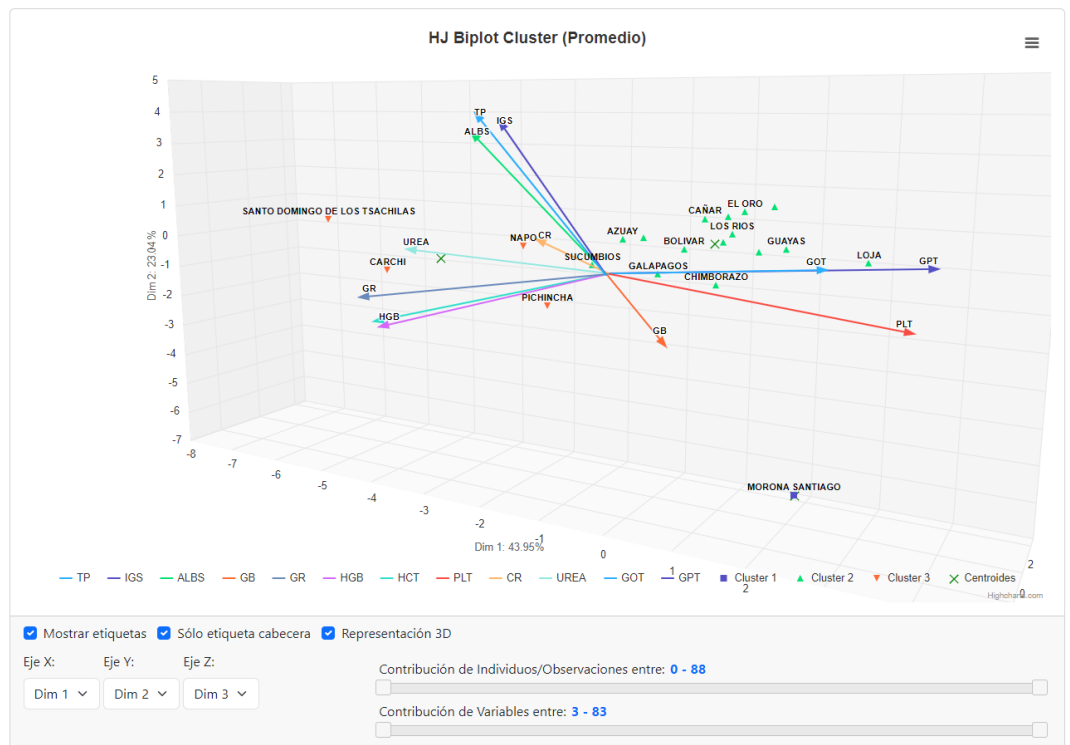


Figura 4.7. Clúster por Lugar de Nacimiento en HJ Biplot Interactivo y Dinámico.

Fuente: Elaboración Propia.

En la Figura 4.8 del Clúster en HJ Biplot interactivo y dinámico, se exponen las siguientes observaciones:

- Las variables Glóbulos Rojos (GR), Hemoglobulina (HGB) y Hematocrito (HCT) tienen una fuerte correlación directa positiva. Se entiende que si aumentan los valores de Glóbulos Rojos (GR), también aumentarán los valores de laboratorio clínico de Hemoglobulina (HGB) y Hematocrito (HCT).
- Los pacientes con diagnóstico patológico C73 - TUMOR MALIGNO DE LA GLÁNDULA TIROIDES, C186 - TUMOR MALIGNO DEL COLON DESCENDENTE, C187 - TUMOR MALIGNO DEL COLON SIGMOIDE, C50 - TUMOR MALIGNO DE LA MAMA, C18 - TUMOR MALIGNO DEL COLON, C501 - TUMOR MALIGNO DE LA PORCIÓN CENTRAL DE LA MAMA y C34 - TUMOR MALIGNO DE LOS BRONQUIOS Y DEL PULMÓN tienen un comportamiento similar en relación con los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023.
- Los pacientes con diagnóstico patológico C182 - TUMOR MALIGNO DEL COLON ASCENDENTE, C829 - LINFOMA NO HODGKIN FOLICULAR, SIN OTRA ESPECIFICACIÓN, C67 - TUMOR MALIGNO DE LA VEJIGA URINARIA, C81 - ENFERMEDAD DE

HODGKIN y C17 - TUMOR MALIGNO DEL INTESTINO DELGADO tienen una preferencia en relación con los resultados de las pruebas de laboratorio clínico Glóbulos Blancos (GB) sobre el periodo 2019-2023.

- Los pacientes con diagnóstico patológico D27 - TUMOR BENIGNO DEL OVARIO y C221 - CARCINOMA DE VÍAS BILIARES INTRAHEPÁTICAS conforman el primer conglomerado (clúster). Tienen un comportamiento diferente al resto de diagnósticos con los resultados de las pruebas de laboratorio clínico sobre el periodo 2019-2023.

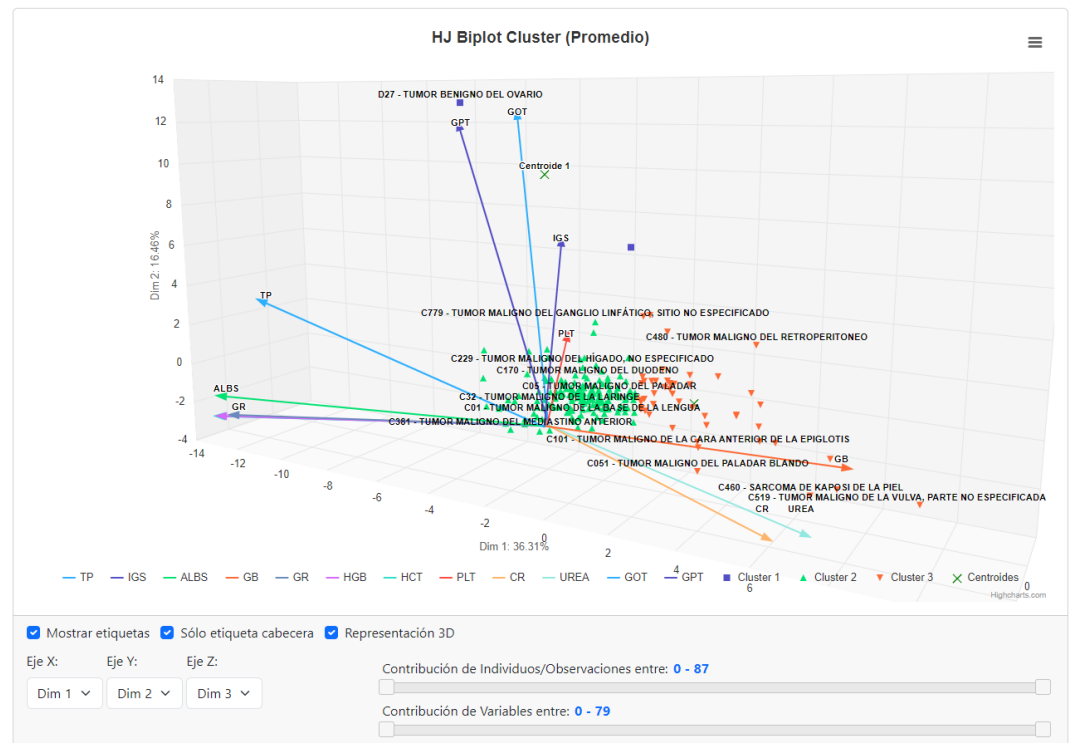


Figura 4.8. Clúster por Diagnóstico en HJ Biplot Interactivo y Dinámico.

Fuente: Elaboración Propia.

4.3.3 ANÁLISIS DE COMPONENTES PRINCIPALES SPARSE (SPCA)

Por medio de la aplicación web interactiva, se genera la matriz de correlación de las variables del conjunto de datos que se está analizando, como se aprecia en la Figura 4.9, la misma que se la puede descargar en un archivo con formato Excel.

Archivo Gráficas Resumen **Matriz Correlación**

[Descargar Archivo Excel](#)

	TP	IGS	ALBS	GB	GR	HGB	HCT
TP	1	0.66194973526614	0.719848748528382	-0.194821897095035	0.390454750090161	0.403721928329517	0.425922104962
IGS	0.66194973526614	1	-0.0415750592871104	0.09025272753492081	0.0088008204601675	-0.0422137296639198	-0.01005080970
ALBS	0.719848748528382	-0.0415750592871104	1	-0.340366576008631	0.511070668294339	0.577639414206037	0.576772529378
GB	-0.194821897095035	0.09025272753492081	-0.340366576008631	1	-0.0738010172740985	-0.151651273465371	-0.14397397876
GR	0.390454750090161	0.0088008204601675	0.511070668294339	-0.0738010172740985	1	0.873193540492724	0.906721459023
HGB	0.403721928329517	-0.0422137296639198	0.577639414206037	-0.151651273465371	0.873193540492724	1	0.981176974565988
HCT	0.425922104962141	-0.010050809709046	0.576772529378229	-0.143973978762666	0.906721459023023	0.981176974565988	1
PLT	0.0603921571167249	0.169308414431066	-0.0768229528169288	0.205924471376643	0.0265743442668513	-0.138582776722836	-0.09432588290
CR	-0.0154964937266288	0.0520996866790957	-0.0672086419124056	0.16287295951841	-0.148512122609825	-0.107585358146712	-0.11970022510
UREA	-0.174635831866099	0.008297523022996221	-0.237420078384733	0.241064485299938	-0.229055183500403	-0.203824116246778	-0.21213643371
GOT	-0.00747670874155914	0.0925205370591644	-0.0937264101351218	0.0385690283358916	-0.0383386408642744	-0.021126886595476	-0.03285263937
GPT	0.0421100488719994	0.0589751991668169	0.00330229674951745	0.0143305167021437	0.0609240683933066	0.08107583398843481	0.072863483173

Figura 4.9. Matriz de Correlación de Variables.

Fuente: Elaboración Propia.

Se realiza un análisis de la correlación entre las variables de las pruebas de laboratorio clínico sobre el periodo 2019-2023, en pacientes con patología diagnóstica de cáncer.

Tabla 1. Matriz de Correlaciones de Variables

LABOR. CLINICO	TP	IGS	ALBS	GB	GR	HGB	HCT	PLT	CR	UREA	GOT	GPT
TP	1,000	0,662	0,720	-0,195	0,390	0,404	0,426	0,061	-0,016	-0,175	-0,008	0,042
IGS	0,662	1,000	-0,042	0,090	0,009	-0,042	-0,010	0,169	0,052	0,008	0,092	0,059
ALBS	0,720	-0,042	1,000	-0,340	0,511	0,578	0,577	-0,077	-0,067	-0,237	-0,094	0,003
GB	-0,195	0,090	-0,340	1,000	-0,074	-0,152	-0,144	0,206	0,163	0,241	0,039	0,014
GR	0,390	0,009	0,511	-0,074	1,000	0,873	0,907	0,027	-0,148	-0,229	-0,038	0,061
HGB	0,404	-0,042	0,578	-0,152	0,873	1,000	0,981	-0,138	-0,108	-0,204	-0,021	0,081
HCT	0,426	-0,010	0,577	-0,144	0,907	0,981	1,000	-0,094	-0,120	-0,212	-0,033	0,073
PLT	0,061	0,169	-0,077	0,206	0,027	-0,138	-0,094	1,000	-0,054	-0,111	-0,059	-0,064
CR	-0,016	0,052	-0,067	0,163	-0,148	-0,108	-0,120	-0,054	1,000	0,712	-0,012	-0,041
UREA	-0,175	0,008	-0,237	0,241	-0,229	-0,204	-0,212	-0,111	0,712	1,000	0,083	-0,007
GOT	-0,008	0,092	-0,094	0,039	-0,038	-0,021	-0,033	-0,059	-0,012	0,083	1,000	0,737
GPT	0,042	0,059	0,003	0,014	0,061	0,081	0,073	-0,064	-0,041	-0,007	0,737	1,000

Fuente: Elaboración Propia.

De acuerdo con la Tabla 1, se muestra la correlación entre cada par de variables, donde se puede observar que en 3 pares existe una asociación mayor de 0.8 lo que podría entenderse como una relación lineal alta entre las variables implicadas, en 7 pares existe una asociación mayor de 0.5 pero menor de 0.8 lo que podría entenderse como una relación lineal media entre las variables implicadas, en 7 pares se da una asociación mayor de 0.1 pero menor de 0.5 lo que podría entenderse como una relación lineal baja entre las variables implicadas y en los 49 pares restantes se detectó una asociación casi nula con una relación lineal menor de 0.1.

El Análisis de Componentes Principales Sparse, obtuvo el 82.9% de varianza explicada, como se ilustra en la Figura 4.10.

	Componente1	Componente2	Componente3	Componente4	Componente5	Componente6
% Varianza	25.5447	14.4881	13.8159	12.5683	8.9149	7.5772
% Varianza Acumulada	25.5447	40.0328	53.8487	66.417	75.3319	82.9091

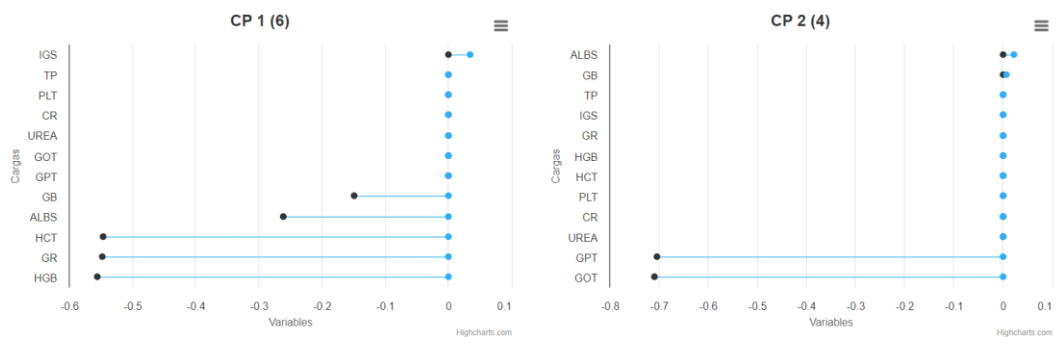
Figura 4.10. Porcentaje de varianza geométrica explicada (ACP Sparse).

Nota: Obtenido a partir de la especificación de cargas no nulas por cada componente (6 componentes).

Fuente: Elaboración Propia.

En la Figura 4.11, se aprecia que los componentes 1,2,3,4,5,6, tienen 6,4,3,2,3,2 valores de carga no nulos; y se evidencian las variables que cargan en cada componente.

Se puede interpretar que las variables Hemoglobulina (HGB), Glóbulos Rojos (GR) y Hematocrito (HCT) son las que se representan mejor en el componente 1, mientras que Glutamato Oxalacetato Transaminasa (GOT) y Glutamato Piruvato Transaminasa (GPT) se representan mejor en el componente 2, con esta misma lógica se interpretan los demás constructos y sus variables asociadas.



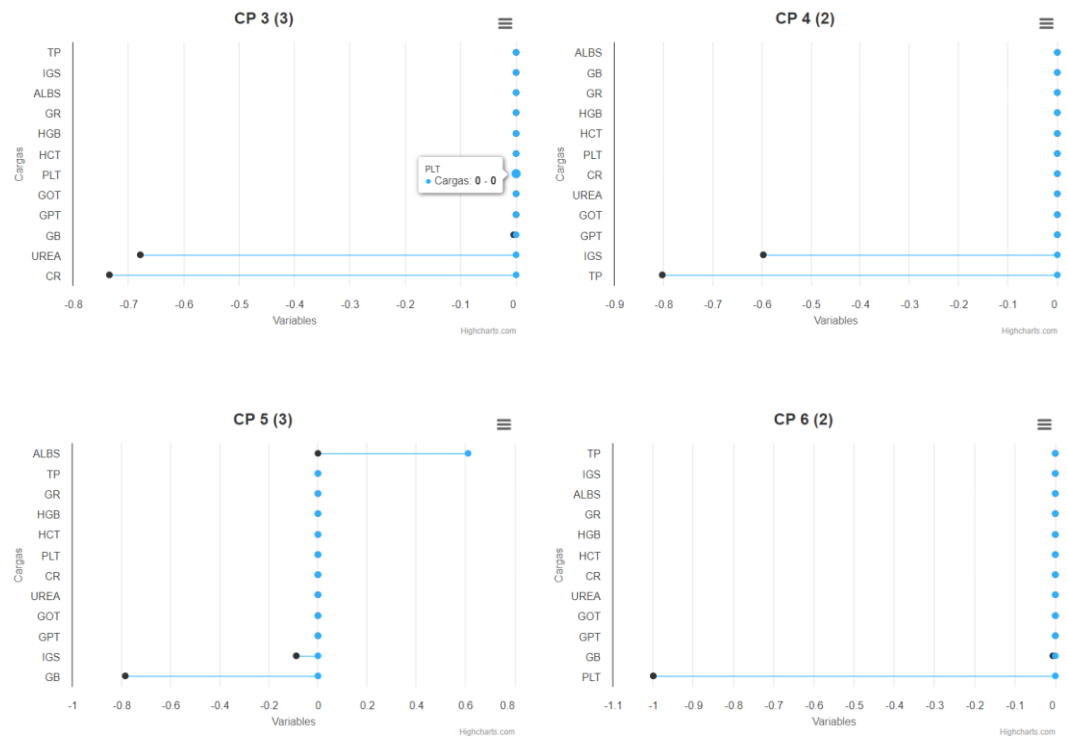


Figura 4.11. Análisis de Componentes Principales Sparse (SPCA).

Nota: Obtenido a partir de la especificación de cargas no nulas por cada componente (6 componentes).

Fuente: Elaboración Propia.

El Análisis de Componentes Principales Sparse aplicando penalización Lasso por cada componente (0.01, 0.16, 0.07, 0.02, 0.5, 0.5), obtuvo el 82.11% de varianza explicada, como se ilustra en la Figura 4.12.

Se puede apreciar que la varianza explicada por las 6 componentes es mayor que en el caso anterior.

	Componente1	Componente2	Componente3	Componente4	Componente5	Componente6
% Varianza	27.1698	14.4724	13.6645	11.636	7.6929	7.4823
% Varianza Acumulada	27.1698	41.6422	55.3067	66.9427	74.6356	82.1179

Figura 4.12. Porcentaje de varianza geométrica explicada (ACP Sparse) con penalización Lasso (lambda).

Nota: Obtenido a partir de la especificación de penalización Lasso (lambda) por cada componente (6 componentes).

Fuente: Elaboración Propia.

En la Figura 4.13, se ilustra que los componentes 1,2,3,4,5,6, tienen (0.01, 0.16, 0.07, 0.02, 0.5, 0.5) valores de penalización Lasso (lambda); y se evidencian las variables que cargan en cada componente.

Se puede interpretar que las variables Hemoglobulina (HGB), Hematocrito (HCT), Glóbulos Rojos (GR) y Albumina Sérica (ALBS) son las que se representan mejor en el componente 1, mientras que Glutamato Oxalacetato Transaminasa (GOT) y Glutamato Piruvato Transaminasa (GPT) se representan mejor en el componente 2, con esta misma lógica se interpretan los demás constructos y sus variables asociadas.

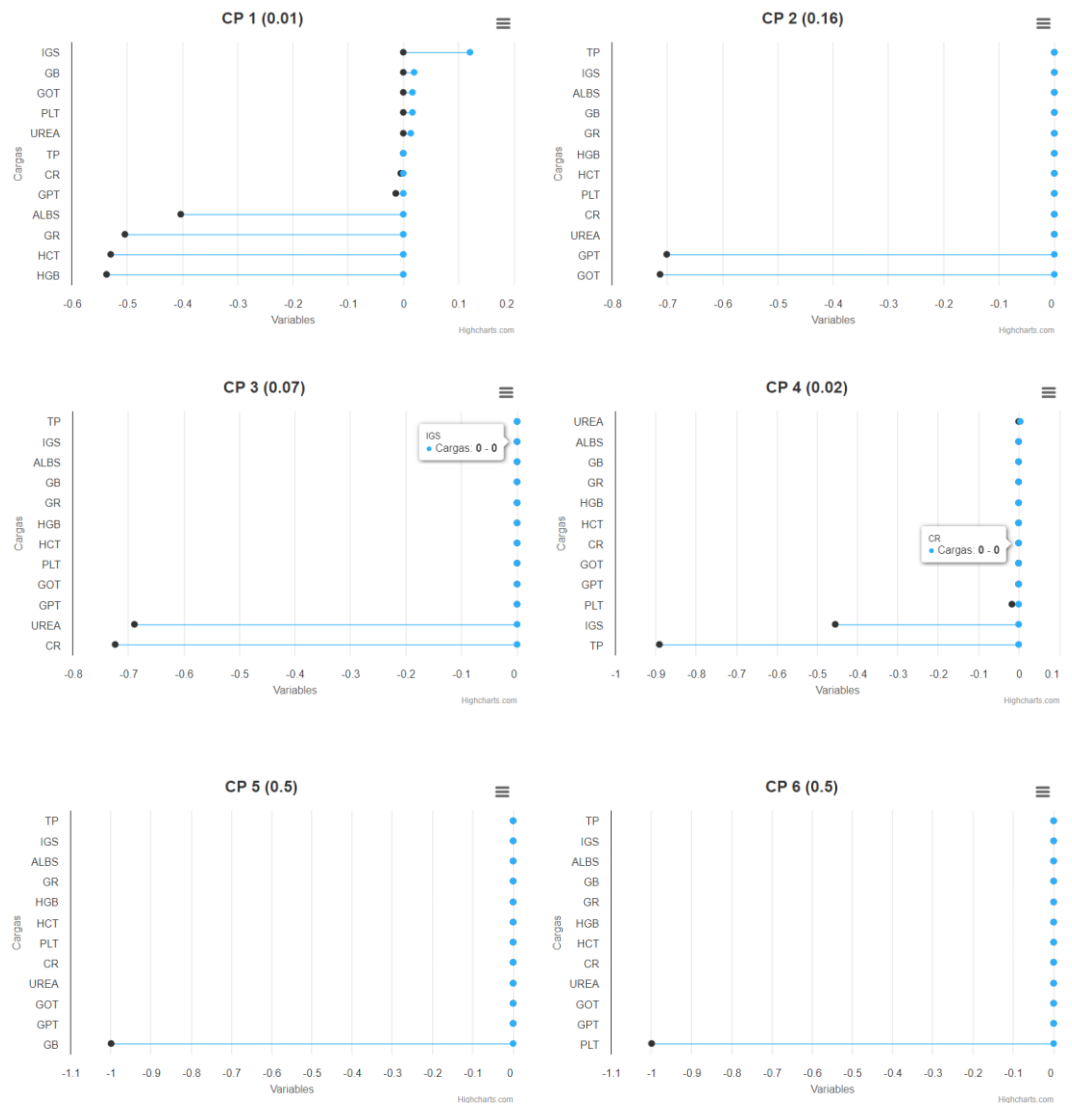


Figura 4.13. Análisis de Componentes Principales Sparse (SPCA) con penalización Lasso (lambda).

Nota: Obtenido a partir de la especificación de penalización Lasso (lambda) por cada componente (6 componentes).

Fuente: Elaboración Propia.

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

1. La revisión de la bibliografía científica expone de manifiesto que es posible la integración entre la programación web con los métodos estadísticos multivariantes como lo es el HJ Biplot.
2. El diseño de la interfaz web interactiva facilita el análisis estadístico multivariante por medio de la aplicación de clúster en un HJ Biplot.
3. Es factible analizar datos masivos de un Hospital Oncológico como SOLCA, que almacena pruebas de laboratorio clínico, extrayendo sus datos en el periodo comprendido entre el año 2019 al 2023 sobre pacientes con patología diagnóstica de cáncer, atendidos desde la Unidad de Pre-Admisión.
4. La aplicación web interactiva, representa gráficos coherentes acorde al criterio médico facilitando la toma de decisiones.

RECOMENDACIONES

1. Implementar nuevos métodos de la familia Biplot tomando como base la aplicación web interactiva desarrollada.
2. Expandir el desarrollo de nuevos algoritmos y técnicas de machine learning en otros métodos tales como la familia Statis, sobre aplicaciones web interactivas.
3. Difundir y promover el desarrollo de aplicaciones web interactivas para expandir el conocimiento a futuras generaciones de especialistas en análisis multivariantes.

6. REFERENCIAS

- [1] Galindo-Villardón, P. (1986). Una alternativa de representación simultánea: HJ-Biplot. *Qüestiió*, Vol. 10 # 1, 13-23.
- [2] Nieto, A. B., Galindo, M. P., Leiva, V., & Vicente-Galindo, P. (2014). A methodology for biplots based on bootstrapping with R. *Revista colombiana de estadística*, 37(2), 367-397.
- [3] Gabriel, K. R. (1971). The biplot graphic display of matrices with application to Principal component analysis. *Biometrika*, 58(3), 453-467.
- [4] Nieto-Librero, A. B., Sierra, C., Vicente-Galindo, M. P., Ruíz-Barzola, O., & Galindo-Villardón, M. P. (2017). Clustering Disjoint HJ-Biplot: A new tool for identifying pollution patterns in geochemical studies. *Chemosphere*, 176, 389-396.
- [5] Cubilla-Montilla, M., Nieto-Librero, A. B., Galindo-Villardón, M. P., & Torres-Cubilla, C. A. (2021). Sparse HJ biplot: A new methodology via elastic net. *Mathematics*, 9(11), 1298.
- [6] Hernández Suárez, M., Molina Pérez, D., Rodríguez-Rodríguez, E. M., Díaz Romero, C., Espinosa Borreguero, F., & Galindo-Villardón, P. (2016). The compositional HJ-biplot—A new approach to identifying the links among bioactive compounds of tomatoes. *International Journal of Molecular Sciences*, 17(11), 1828.
- [7] González, L. V. (2023). Generalización del biplot logístico para dos o más matrices (Doctoral dissertation, Universidad de Salamanca).
- [8] Gabriel, K. R. y Odoroff, C. L. (1990). Biplots in biomedical research. *Statistics in medicine*, 9(5):469–485.
- [9] Gower, J. C. y Hand, D. J. (1995). *Biplots*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- [10] IBM Corp. (2024). *Ibm spss statistics*.
- [11] SAS Institute Inc. (2024). *Statistical analysis software (sas)*.
- [12] LLC (2024). *Minitab 21*.
- [13] StataCorp (2024). *Stata statistical software*
- [14] Yan, W., Kang, M. S., Ma, B., Woods, S., & Cornelius, P. L. (2007). GGE biplot vs. AMMI analysis of genotype-by-environment data. *Crop science*, 47(2), 643-653.
- [15] Kovach, W. (1999). *Multivariate statistical package (mvsp)*
- [16] Inc. Statgraphics Technologies (2024). *Statgraphics*
- [17] James Grace, Martin Hutten, B. M. M. J. P. y Hatch, W. (2018). *Pc-ord*
- [18] ter Braak, C. J., & Šmilauer, P. (2012). *Canoco reference manual and user's guide: software for ordination, version 5.0*.
- [19] Ltd. Analyse-it Software (2024). *Analyse-it*.
- [20] Udina, F. (2004). Interactive biplot construction. *Journal of Statistical Software*, 13.
- [21] Udina, F. (2005). *Xls-biplot*.
- [22] Young, F. W. (1990). *Vista the visual statistics system*

- [23] Highland Statistics Ltd. (2017). Brodgar
- [24] Vicente-Villardón, J. L. (2024). Multiblot
- [25] R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
- [26] Libroero, A. B. N., Baccala, N., Galindo, P. V., y Villardón, P. G. (2022). multiblotGUI: Multiblot Analysis in R. R package version 1.1
- [27] Libroero, A. B. N., Villardón, P. G., y Freitas, A. (2019). biplotbootGUI: Bootstrap on Classical Biplots and Clustering Disjoint Biplot. R package version 1.2
- [28] Dumble, S. (2022). GGEBiplots: GGE Biplots with 'ggplot2'. R package version 0.1.3
- [29] Dray, S., Dufour, A.-B., Thioulouse, J., Jombart, T., Pavoine, S., Lobry, J. R., Ollier, S., Borcard, D., Legendre, P., Bougeard, S., y Siberchicot, A. (2022). Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences. R package version 1.7-19
- [30] Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., y Wagner, H. (2017). vegan: Community Ecology Package. R package version 2.4-5
- [31] Babativa-Márquez, J. G. (2020). BiplotML: Biplots Estimation with Machine Learning Algorithms
- [32] Vicente-Villardón, J. L. (2021). MultBiplotR: Multivariate Analysis Using Biplots in R. R package version 1.6.14
- [33] Cubilla Montilla, M. I. (2019). Contribuciones al análisis biplot basadas en soluciones factoriales disjuntas y en soluciones sparse.
- [34] Winter, A., Haux, R., Ammenwerth, E., Brigl, B., Hellrung, N., Jahn, F., ... & Jahn, F. (2011). Health information systems (pp. 33-42). Springer London.
- [35] Ávila Arancibia, E. (2016), Sistemas de información hospitalaria (HIS). Disponible en: <https://inglomayor.cl/edicion11/Works/Eduardo%20Avila/HIS-Eduardo%20%C3%81vila%20Arancibia.pdf>.
- [36] Vargas-Herrera, J., Segovia-Juarez, J., & GM, G. N. (2015). Information system of the national network of public health laboratories in Peru (Netlab). Revista Peruana de Medicina Experimental y Salud Pública, 32(2), 378-384.
- [37] De León, A., & Gloria, A. (2013). Utilización inadecuada de los avances científicos técnicos del laboratorio clínico y del método clínico. Repercusión en los servicios de salud. Revista Médica Electrónica, 35(4), 386-396.
- [38] Sepulveda, J. L., & Young, D. S. (2013). The ideal laboratory information system. Archives of Pathology and Laboratory Medicine, 137(8), 1129-1140.
- [39] López, A. (2012). Sistemas de información del laboratorio clínico. Secretario editorial David Rojas Comité editorial, 109.
- [40] Paredes Bravo, M. A. (2023). Análisis de comparación de los lenguajes de programación en Php y Python utilizada para el desarrollo de páginas web (Bachelor's thesis, Babahoyo: UTB-FAFI. 2023).
- [41] PHP (2024). Hypertext Preprocessor (PHP). [En línea]. Disponible en: <https://www.php.net/manual/es/intro-whatish.php>.

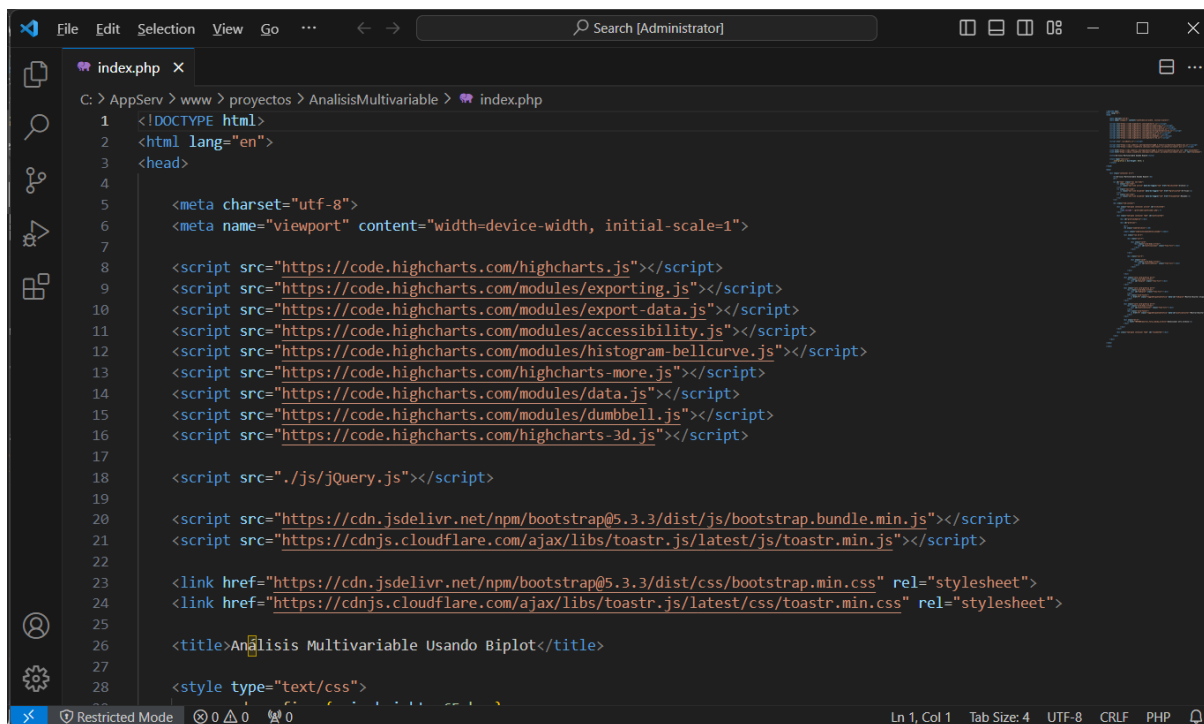
- [42] JAVA (2024). Java (Lenguaje de programación). [En línea]. Disponible en: https://www.java.com/en/download/help/whatis_java.html.
- [43] Amazon Web Services (AWS) (2024). Java (Lenguaje de programación). [En línea]. Disponible en: <https://aws.amazon.com/es/what-is/java/>.
- [44] Amazon Web Services (AWS) (2024). Python (Lenguaje de programación). [En línea]. Disponible en: <https://aws.amazon.com/es/what-is/python/>.
- [45] TIOBE (2024). Python (Lenguaje de programación). [En línea]. Disponible en: <https://www.tiobe.com/tiobe-index/>.
- [46] NetBeans (2024). NetBeans IDE (Entorno de Desarrollo Integrado). [En línea]. Disponible en: <https://www.oracle.com/es/tools/technologies/netbeans-ide.html>.
- [47] Eclipse (2024). Eclipse IDE (Entorno de Desarrollo Integrado). [En línea]. Disponible en: <https://eclipseide.org/>.
- [48] WordPress (2024). WordPress CMS (Sistema de Gestión de Contenidos). [En línea]. Disponible en: <https://wordpress.com/>.
- [49] Bootstrap (2024). Bootstrap (front-end framework). [En línea]. Disponible en: <https://getbootstrap.com/>.
- [50] FusionCharts (2024). FusionCharts (Biblioteca). [En línea]. Disponible en: <https://www.fusioncharts.com/>.
- [51] Chart.js (2024). Chart.js (Biblioteca). [En línea]. Disponible en: <https://www.chartjs.org/>.
- [52] Highcharts (2024). Highcharts (Biblioteca). [En línea]. Disponible en: <https://www.highcharts.com/>.
- [53] JavaScript (2024). JavaScript (Biblioteca). [En línea]. Disponible en: <https://developer.mozilla.org/es/docs/Web/JavaScript>.
- [54] jQuery (2024). jQuery (Biblioteca). [En línea]. Disponible en: <https://jquery.com/>.

7. GLOSARIO

JavaScript Object Notation (JSON), Es un formato de texto ligero para el intercambio de datos. JSON es un subconjunto de la notación literal de objetos de JavaScript. Sin embargo, en la actualidad debido a su amplia adopción como alternativa a XML, se considera un formato de lenguaje independiente.

Laboratorio Clínico: Es el área física, en donde tanto los técnicos como los profesionales en análisis clínicos, realizan el análisis de las muestras biológicas humanas que contribuyen al estudio, prevención, detección, diagnóstico, tratamiento y seguimiento de las enfermedades.

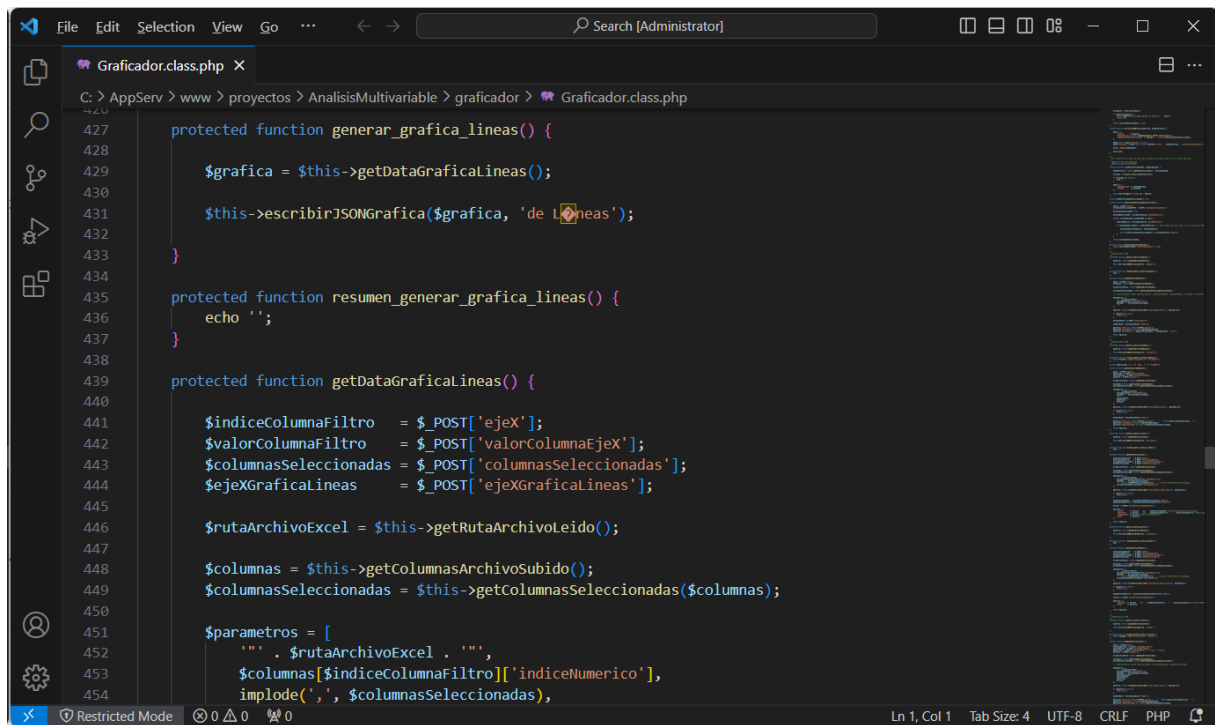
8. APÉNDICES Y ANEXOS



```
1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4
5     <meta charset="utf-8">
6     <meta name="viewport" content="width=device-width, initial-scale=1">
7
8     <script src="https://code.highcharts.com/highcharts.js"></script>
9     <script src="https://code.highcharts.com/modules/exporting.js"></script>
10    <script src="https://code.highcharts.com/modules/export-data.js"></script>
11    <script src="https://code.highcharts.com/modules/accessibility.js"></script>
12    <script src="https://code.highcharts.com/modules/histogram-bellcurve.js"></script>
13    <script src="https://code.highcharts.com/highcharts-more.js"></script>
14    <script src="https://code.highcharts.com/modules/data.js"></script>
15    <script src="https://code.highcharts.com/modules/dumbbell.js"></script>
16    <script src="https://code.highcharts.com/highcharts-3d.js"></script>
17
18    <script src="./js/jquery.js"></script>
19
20    <script src="https://cdn.jsdelivr.net/npm/bootstrap@5.3.3/dist/js/bootstrap.bundle.min.js"></script>
21    <script src="https://cdn.jsdelivr.net/npm/toastr.js/latest/js/toastr.min.js"></script>
22
23    <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.3.3/dist/css/bootstrap.min.css" rel="stylesheet">
24    <link href="https://cdn.jsdelivr.net/npm/toastr.js/latest/css/toastr.min.css" rel="stylesheet">
25
26    <title>Análisis Multivariable Usando Biplot</title>
27
28    <style type="text/css">
```

Figura A1. Código de Programa PHP para Interfaz Web de Visualización.

Fuente: Elaboración Propia.



```
427 protected function generar_grafica_lineas() {
428
429     $grafica = $this->getDataGraficaLineas();
430
431     $this->escribirJSONGrafica($grafica, 'de Líneas');
432
433 }
434
435 protected function resumen_generar_grafica_lineas() {
436     echo '';
437 }
438
439 protected function getDataGraficaLineas() {
440
441     $indiceColumnaFiltro = $_POST['ejeX'];
442     $valorColumnaFiltro = $_POST['valorColumnaEjeX'];
443     $columnasSeleccionadas = $_POST['columnasSeleccionadas'];
444     $ejeXGraficaLineas = $_POST['ejeXGraficaLineas'];
445
446     $rutaArchivoExcel = $this->getRutaArchivoLeído();
447
448     $columnas = $this->getColumnasArchivoSubido();
449     $columnasSeleccionadas = $this->getColumnasSeleccionadas($columnas);
450
451     $parametros = [
452         '' . $rutaArchivoExcel . '',
453         $columnas[$indiceColumnaFiltro]['indiceNumerico'],
454         implode(',', $columnasSeleccionadas),
```

Figura A2. Código de programa PHP para Gráfico Interactivo y Dinámico.

Fuente: Elaboración Propia.

```
File Edit Selection View Go ... Search [Administrator]
formSubidaArchivo.php x
C: > AppServ > www > proyectos > AnalisisMultivariable > graficador > formSubidaArchivo.php
1 <?php
2 /**
3  * Formulario de subida del archivo Excel/csv/txt
4  *
5  * @var string $urlAccion
6  * @var string $metodo
7  * @var string $errores
8  *
9  */
10 ?><form action="<?php echo $urlAccion ?>" method="post" enctype="multipart/form-data">
11
12     <input type="hidden" name="METODO" value="<?php echo $metodo ?>" />
13
14     <h4>Selección del archivo</h4>
15
16     <?php if (!empty($errores)) { ?>
17         <div class="alert alert-danger"><?php echo $errores ?></div>
18     <?php } ?>
19
20     <p></p>
21
22     <div class="card">
23         <div class="card-body">
24
25             <div class="mb-2">
26                 <label for="archivo" class="form-label">Selecione el archivo:</label>
27                 <input type="file" class="form-control" id="archivo" placeholder="Selecione el archivo" name="AR
28             </div>
29
30         </div>
31     </div>
32
33 </form>
34
35 </?php
```

Figura A3. Código de programa PHP para Gestión de Archivos.

Fuente: Elaboración Propia.

```
8 # Todos los parámetros son requeridos
9
10 rutaArchivo      <- args[1]
11 indiceColumnaFiltro <- as.numeric(args[2])
12 columnasSeleccionadas <- str_split(args[3], ",", simplify = TRUE)
13 filtro           <- args[4]
14 indiceColumnaEjeX  <- as.numeric(args[5])
15
16 datos <- openxlsx::read.xlsx(rutaArchivo)
17
18 # Filtrar los datos de acuerdo con la columna 'indiceColumnaFiltro'
19
20 if (filtro != "-1") {
21   condicion <- c(colnames(datos)[indiceColumnaFiltro])
22   datos <- filter(datos, .data[[condicion[1]]] == filtro)
23 }
24
25 # Armar data.frame sólo con las columnas seleccionadas
26
27 columnas = data.frame(matrix(ncol = 0, nrow = nrow(datos)));
28
29 for(indiceColumnaStr in columnasSeleccionadas) {
30   indiceColumna <- as.numeric(indiceColumnaStr)
31   columnas <- cbind(columnas, datos[indiceColumna])
32 }
33
34 promedios <- colMeans(columnas, na.rm = TRUE)
35
36 toJSON(
37   list(
38     promedios = as.list(promedios)
39   )
40 )
41
42 )
```

Figura A4. Código de programa R para Gráfico Exploratorio Circular.

Fuente: Elaboración Propia.

```
54 # VARNUM
55 # Componentes principales sparse obtenidos a partir de la especificación de cargas no nulas por componente
56 sparseAnálisisVarnum <- spca(datos, K = componentesDeseados, type = "Gram", sparse = "varnum", para = paraVarnum)
57 cargasVarnum <- sparseAnálisisVarnum$loadings
58
59 porcentajeVarianzaVarnum <- t(data.frame(round((sparseAnálisisVarnum$pev) * 100, 4)))
60 colnames(porcentajeVarianzaVarnum) <- paste0('Componente', 1:componentesDeseados)
61 rownames(porcentajeVarianzaVarnum) <- c('')
62
63 porcentajeVarianzaVarnumAcum <- t(data.frame(round((sparseAnálisisVarnum$pev) * 100, 4)))
64 porcentajeVarianzaVarnumAcum <- round(cumsum(porcentajeVarianzaVarnumAcum), 4)
65
66 tablaVarianzasVarnum <- rbind(porcentajeVarianzaVarnum, porcentajeVarianzaVarnumAcum)
67 rownames(tablaVarianzasVarnum) <- c('% Varianza', '% Varianza Acumulada')
68
69 # PENALTY
70 # Componentes principales sparse obtenidos a partir de la especificación del lambda 2
71 sparseAnálisis <- spca(datos, K = componentesDeseados, type = "Gram", sparse = "penalty", trace = FALSE, para = paraPenalty)
72 cargasPenalty <- sparseAnálisis$loadings
73
74 porcentajeVarianza <- t(data.frame(round((sparseAnálisis$pev) * 100, 4)))
75 colnames(porcentajeVarianza) <- paste0('Componente', 1:componentesDeseados)
76 rownames(porcentajeVarianza) <- c('')
77
78 porcentajeVarianzaAcum <- t(data.frame(round((sparseAnálisis$pev) * 100, 4)))
79 porcentajeVarianzaAcum <- round(cumsum(porcentajeVarianzaAcum), 4)
80
81 tablaVarianzas <- rbind(porcentajeVarianza, porcentajeVarianzaAcum)
82 rownames(tablaVarianzas) <- c('% Varianza', '% Varianza Acumulada')
83
84 cargasVarnumDF <- as.data.frame(cargasVarnum)
85 cargasPenaltyDF <- as.data.frame(cargasPenalty)
86
87 listaCargasVarnum <- list()
88 listaCargasPenalty <- list()
```

Figura A5. Código de programa R para Sparse PCA.

Fuente: Elaboración Propia.

El código fuente de la aplicación web interactiva, se encuentra alojado en un repositorio de la nube, al que se puede acceder por medio del enlace:

https://espolec-my.sharepoint.com/personal/serabauz_espolec/layouts/15/onedrive.aspx?e=5%3Af4a9a80a5f314cdcb2b470210886be20&sharingv2=true&fromShare=true&at=9&CT=1716435516578&OR=OWA%2DNT%2DMail&CID=02e9f83b%2D40e1%2D47ce%2Dc725%2D56df7e5e6cfb&FolderCTID=0x01200088D7F38F9A2ADA48BC635897D22674F7&id=%2Fpersonal%2Fserabauz%5Fespolec%5Fedu%5Fec%2FDocuments%2FMESAP%2FCristian%20Tomala