

UNIVERSIDAD PERUANA UNIÓN

Facultad de Ingeniería y Arquitectura

E.A.P. INGENIERÍA DE SISTEMAS



TESIS DE TITULACION

Tesis presentada para optar el título profesional de Ingeniero de Sistemas

Aspecto Nutricional apoyado por un Modelo Bayesiano y su relación en el pronóstico de Morbilidad del Neonato de la gestante.

Bach. Artica Chacón, Miguel Ángel

Bach. Ocaña Sudario, Néstor Anderson

Asesor:

Dr. Juan Jesús Soria Quijaite

Lima, Diciembre del 2015

Agradecimiento

En primer lugar A Dios por la vida y por las bendiciones que nos ha dado. Todo lo que somos y lo que sabemos es por Él, porque Él lo permite.

Un agradecimiento a nuestros familiares especialmente a nuestros padres por el apoyo y amor que siempre nos demostraron; así como su comprensión por las horas no compartidas.

A la Universidad Peruana Unión y a todos sus representantes, que hacen posible que nos sigamos formando.

Finalmente, agradecemos a nuestro asesor Dr. Juan Jesús Soria Quijaite, por los esfuerzos y orientaciones que hicieron posible culminar satisfactoriamente este proyecto de investigación.

Dedicatoria

Dedicamos este trabajo al creador de todas las cosas, por la vida y por permitir haber llegado a este momento tan importante de nuestra formación profesional; el que nos ha dado la fortaleza para continuar, con toda humildad este trabajo está dedicado primeramente a Dios.

Al ser más querido, a nuestras madres, Clara y Máxima, por demostrarnos su cariño y apoyo incondicional.

Al apoyo incondicional de nuestros hermanos, los más cercanos críticos y amigos al mismo tiempo.

En memoria de nuestro amigo y colega, Fredy Aguilar Huaranga, aunque no está presente, sabemos que disfrutaría al igual que nosotros del logro obtenido.

INDICE

RESUMEN	10
ABSTRACT	11
Introducción	12
Capítulo I:	14
Planteamiento metodológico:	14
1.1. Título de la investigación.	14
1.2 Problema objeto de investigación	14
1.2.1 Planteamiento del problema	14
1.2.2 Formulación del problema	17
A. Problema general.....	17
B. Problemas específicos.	17
1.3. Objetivos:	19
1.3.1 Objetivo General.	19
1.3.2 Objetivos específicos.....	19
1.4. Hipótesis:	21
1.4.1 Hipótesis general.....	21
1.4.2 Hipótesis específicas.	21
1.5 Operacionalización de variables:	23
1.5.1 Variables dependiente e independiente.	23
1.5.2 Definición operacional de las variables.....	23
1.6 Matriz de Consistencia	27
1.7 Justificación:	35
1.8 Alcances de la investigación.	35
1.9 Limitaciones de la investigación.	35
Capítulo II:	37
Marco Teórico - Conceptual	37
2.1 Estado del arte (antecedentes)	37
2.2 Marco teórico.	41
2.3 Marco conceptual.	51
Capítulo III:	67
Métodos y Materiales	67
3.1 Metodología de Investigación.	67
3.1.1 Nivel de investigación.....	67
3.1.2 Tipo de investigación.	68
3.1.3 Enfoque de la investigación.	68

3.1.4.	Método de intervención tecnológica	68
3.1.5	Población y muestra	69
3.1.6	Presentación de la información.	70
3.2	Materiales:	71
3.3	Cronograma de actividades.	72
3.4	Costos de la investigación.	74
CAPÍTULO IV:		76
INGENIERÍA DE LA PROPUESTA		76
4.1	Proceso de pronóstico de la gestante	76
4.2	Diagramas De Interfaz de la Ingeniería	78
4.3	Restricciones del Modelo Bayesiano	80
4.4	VARIABLES del Modelo	80
4.5	Topología de la Red Bayesiana	84
4.6	Modelo probabilístico	87
4.7	Descripción de los escenarios de pronóstico.	125
CAPÍTULO V:		126
RESULTADOS DE LA INVESTIGACIÓN		126
5.1	Resultados de los escenarios pronóstico	126
CONCLUSIONES		140
RECOMENDACIONES		145
REFERENCIAS BIBLIOGRÁFICAS		146
ANEXOS		148
Anexo 1. Proceso de morbilidad materna y neonatal		148
Anexo 2. Cronograma de Actividades		149
Anexo 3. Codificación Algoritmo de clasificación		150
Anexo 4. Algoritmo de obtención de Morbilidad		152
Anexo 5. Tabla comparativa de índice de masa corporal en el Perú.		153
Anexo 6. Datos obtenidos de una madre gestante menor de edad		156

Índice de Figuras

Figura 1. Red bayesiana compra de un equipo informático	45
Figura 2. Metodología de Sampieri	67
Figura 3. Diagrama de bloques.....	69
Figura4. Casos de uso del negocio - diseño de Modelo Bayesiano para el pronóstico de morbilidad del neonato	78
Figura 5. Casos de uso del sistema - diseño del Modelo Bayesiano para el pronóstico de morbilidad del neonato	78
Figura6. Diagrama de Colaboración - Autenticación de usuario.	79
Figura7. Diagrama de paquetes	79
Figura 8. Diagrama de morbilidad neonatal	80
Figura 9. Topología de un clasificador Naive Bayes.....	90
Figura 10. Algoritmo K2	93
Figura11. Pseudocodigo del Algoritmo B	94
Figura 12. Pseudocodigo Algoritmo TAN	96
Figura 13. Pseudocodigo Algoritmo EM.....	98
Figura14. Gráfica de la hipótesis aceites y grasas	100
Figura 15. Gráfica de la hipótesis agua	102
Figura 16. Grafica de la hipótesis Azúcar	104
Figura17.Grafica de la hipótesis Carnes y legumbres	106
Figura 18.Gráfica de la hipótesis Cereales	108
Figura 19.Gráfica de la hipótesis complementos vitamínicos.....	110
Figura20.Gráfica de la hipótesis condimentos	112
Figura 21.Gráfica de la hipótesis descalsificantes.....	114
Figura22.Gráfica de la hipótesis frutas.....	116
Figura 23.Gráfica de la hipótesis panes.....	118

Figura 24. Gráfica de la hipótesis productos lácteos	120
Figura 25. Gráfica de la hipótesis sal	122
Figura 26. Gráfica de la hipótesis verduras y hortalizas.....	124
Figura 27. Arquitectura n - capas	128
Figura 28. Dimensiones del sistemas	129
Figura 29. Dimensiones y sus escalas	130
Figura 30. Escala de valores de las dimensiones.....	130
Figura 31. Raciones de alimentos ingeridos de una madre gestante	131
Figura 32. Cálculo de la probabilidad	131
Figura 33. Valor de categoría, según las dimensiones	132
Figura 34. Categoría y sus escalas.....	132
Figura 35. Cálculo de porcentaje de vasos por día.....	133
Figura 36. Valores por día.....	134
Figura 37. Codificación del algoritmo.....	135
Figura 38. Uso del analysis services.....	136
Figura 39. Codificación del algoritmo de clasificación.....	137
Figura 40. Definición de variables	138
Figura 41. Ingreso de datos - Madre gestante.....	139
Figura 42. Entrenamiento del Modelo Bayesiano	139
Figura 43. Proceso de morbilidad materna y neonatal	148
Figura 44 Cronograma de Proyecto.....	149

Índice de Tablas

Tabla 1. Operacionalización de variables.....	23
Tabla 2. Tabla de medidas por raciones	26
Tabla 3. Matriz de consistencia	27
Tabla 4. Tabla de Población y muestra.....	70
Tabla 5. Descripción del Hardware	71
Tabla 6. Cronograma de Actividades	72
Tabla 7. Presupuesto de costos y recursos humanos	74
Tabla 8. Presupuesto de hardware	74
Tabla 9. Presupuesto de costo, software.....	75
Tabla 10. Correlación de la variable aceites y grasas.....	99
Tabla 11. Correlación de la variable agua	101
Tabla 12. Correlación de la variable azúcar	103
Tabla 13. Correlación de la variable carnes y legumbres.....	105
Tabla 14. Correlación de la variable cereal	107
Tabla 15. Correlación de la variable complementos vitamínicos.....	109
Tabla 16. Correlación de la variable condimentos	111
Tabla 17. Correlación de la variable descalcificantes.....	113
Tabla 18. Correlación de la variable frutas.....	115
Tabla 19. Correlación de la variable panes.....	117
Tabla 20. Correlación de la variable productos lácteos.....	119
Tabla 21. Correlación de la variable sal	121
Tabla 22. Correlación de la variable verduras y hortalizas	123
Tabla 23. Tabla de códigos de riesgo	126
Tabla 24. Presupuestos de costos y recursos humanos en el desarrollo	126

Tabla 25. Presupuesto del hardware	127
Tabla 26. Presupuesto de software	127
Tabla 27. Comparativa de IMC en el Perú.	153
Tabla 28. Datos de madre gestante menor de edad	156

RESUMEN

En este trabajo de investigación, al haberse cumplido con los objetivos se concluyó que al usar el modelo bayesiano, basado en algoritmo Naive Bayes, se logró cumplir con el pronóstico de la morbilidad de la madre gestante. Así mismo, usando la interfaz web, se logró integrar los datos de las madres gestantes respecto a su proceso de embarazo, en que el aspecto nutricional de ellas, fue un factor relevante para que el algoritmo bayesiano pronosticara la morbilidad neonatal.

Las variables independientes relacionadas al aspecto nutricional de las madres gestantes y la variable dependiente Morbilidad Neonatal, se relacionaron en forma directa usando el estadístico correlacional del Rho de Spearman en los que se obtuvieron un $P_{\text{valor}} = 0.000$, que por ser menor al nivel de significancia $\alpha=0.05$, se rechazó la hipótesis nula y se aceptó la hipótesis de investigación.

En este estudio tuvimos una muestra de 14 días de una madre gestante que mensualmente asistió a su control de embarazo. La cual se pudo comprobar el estado de morbilidad de sus neonato es medio con un 67 % aproximadamente.

ABSTRACT

In this research, after having met the objectives it is concluded that using the Bayesian model, based on the Naive Bayes algorithm. We were able to know the forecast of morbidity of the pregnant mother, so using the web interface, let's integrate data from pregnant mothers regarding the process of their pregnancy. About of all factor found, the nutritional aspect was a relevant factor to the Bayesian algorithm to predict neonatal morbidity.

The independent variables related to the nutritional aspect of the pregnant women and the dependent variable Neonatal Morbidity are related directly using Spearman's rank correlation coefficient Rho which obtained a result of $P_valor = 0.000$, and for being less than the significance level $\alpha = 0.05$, the null hypothesis is rejected and the research hypothesis is accepted.

This study had a sample of 14 days from a pregnant mother who attended her monthly birth control. Thereby it was found that the state of her neonate morbidity is medium with an approximately 67%.

Introducción

El embarazo, dentro del proceso de la reproducción humana, se desarrolla el aspecto físico, neurológico y mental del neonato, por ello es importante determinar su morbilidad, a fin de lograr una óptima salud en la madres gestante.

En un país multicultural como el Perú, con una diversidad geográfica muy variada, encontramos que los centros públicos de salud están muy alejados de la población, y no están equipados debidamente para un control adecuado de la gestante.

Dentro del aspecto nutricional de la madre gestante y su proceso de embarazo, el control que se debe tener en cuenta es su IMC (índice de masa corporal), porque es muy importante tener el control de su peso de acuerdo a las meses de gestación, pues con estos datos se ayudará al buen desarrollo del feto y su buen estado de salud.

La mayoría de las características mencionadas anteriormente son registradas en diversas formas y formatos, como: cuadernos, software informático, formatos físicos propios del centro de salud, etc.

Así mismo, las redes bayesianas con sus respectivos algoritmos de entrenamiento como son el K2, algoritmo B, algoritmo HC, algoritmo TAN, algoritmo BAN y el algoritmo HC, nos ayudan a relacionar las variables independientes que están evocadas al aspecto nutricional de las madres gestantes con su morbilidad.

Los estudios estadísticos que se realizan con la información recolectada necesitan ser contrastados y complementados con el análisis de otros modelos que respalden el resultado de los mismos.

Los algoritmos de aprendizaje de redes bayesianas se centran en el descubrimiento de relaciones de independencia y relevancia entre variables, por lo que el modelo

resultante refleja de forma explícita numerosas relaciones de interés. En cierto modo, permiten establecer relaciones mucho más ricas que las reglas de asociación.

Además, su explotación posterior mediante el uso de distintas técnicas de inferencia probabilística, nos permitirá ampliar el conjunto de relaciones descubiertas y evaluar las distintas hipótesis que podamos formular.

En cuanto al uso predictivo, nos referimos fundamentalmente al uso de las redes bayesianas como clasificadores. Para esta tarea, veremos que podemos construir redes con distinto nivel de complejidad mediante la adición de restricciones al proceso de aprendizaje.

Dentro de este contexto es necesario implementar una interfaz web, apoyada por la red bayesiana de algoritmo Naive de Bayes que nos permita pronosticar la condición de morbilidad del Neonato a partir de los datos recolectados en el aspecto nutricional de las madres gestantes.

La interfaz con codificación en el lenguaje C# y el uso de *Analysis Services*, permitió medir en forma precisa la morbilidad de la madre gestante bajo el algoritmo de Naive Bayes.

Capítulo I:

Planteamiento metodológico:

1.1. Título de la investigación.

Aspecto nutricional apoyado por un Modelo Bayesiano y su relación en el pronóstico de Morbilidad del Neonato de la gestante.

1.2 Problema objeto de investigación

1.2.1 Planteamiento del problema

El embarazo, dentro del proceso de la reproducción humana, se encuentra dividido en tres trimestres: primer trimestre (semana 1 a semana 12), segundo trimestre (semana 13 a semana 28) y tercer trimestre (semana 29 a semana 40), en la que el proceso de embarazo tiene una duración aproximada de 9 meses.

Los centros públicos de salud en el Perú, no están equipados debidamente para un control adecuado de la morbilidad neonatal de la madre gestante. El aspecto nutricional es un factor muy relevante dentro del proceso del embarazo de la madre gestante, en la que la ausencia y recolección errónea de los datos del aspecto nutricional en cada control por parte del área de neonatología y los especialistas dentro del área, no se suele pronosticar con precisión. Actualmente la inteligencia artificial dentro de la rama de las redes bayesianas, toma un valor muy significativo, porque ayuda a las diferentes áreas como Educación, Medicina Humana, Ingeniería de Sistemas y Psicología, en la toma de decisiones de los pronósticos de morbilidad neonatal.

En la actualidad, en los diferentes hospitales en el área de neonatología, no existe una medición precisa de la morbilidad de la madre

gestante y su neonato, por ello los algoritmos de aprendizaje de redes bayesianas se centran en el descubrimiento de relaciones de independencia y relevancia entre variables, entonces el modelo resultante refleja de forma explícita numerosas relaciones de interés.

No existe un Modelo Bayesiano para el pronóstico del neonato dentro del aspecto nutricional, en el área de neonatología en los diferentes hospitales, con las dimensiones: cereales, verduras y hortalizas; frutas, aceites y grasas; productos lácteos, carnes, legumbres y huevos; complementos, vitamínicos, agua, descalsificantes, IMC (índice de masa corporal), horario de alimentación, panes, complementos dietéticos, con las variables: trigo, harina de maíz, de avena y harina blanca, arroz integral y blanco, sémola, lechuga, espinaca, brócoli, zanahorias, calabazas, papas, maíz, garbanzos, lentejas, plátanos, manzana, naranja, arándanos, frutas secas, pera, kiwi, mangos, uva, melón, piña, aceite de oliva, aceite de maíz, aceite de girasol, aceite de soja, aceite de almendras, manteca, mantequilla, margarina, leches enteras y descremadas, postres, lácteos, quesos, yogures, carne de vaca, carne de cerdo, carne de cordero, mariscos, pavo, pollo, pescados, huevos, otros, ácido fólico, otras vitaminas, agua, cafeína, gaseosa, cigarro, alcohol, alimentos procesados, IMC (índice de masa corporal). Datos como el desayuno, almuerzo y cena para lo cual en la investigación se propone un modelo bayesiano de acorde a la realidad de los diferentes hospitales.

En tal contexto, la ausencia de una interfaz web apoyada por un Modelo Bayesiano, usando el algoritmo Naive de Bayes impide la interpretación de los datos recolectados en forma precisa dentro de las

etapas de la madre gestante, el cual ayuda al pronóstico de la morbilidad del neonato.

Las redes bayesianas es base para el estudio del Modelo Bayesiano, usando el teorema general de probabilidades de Naive Bayes en concordancia con los algoritmos bayesianos, algoritmo K2, algoritmo B, algoritmo HC, algoritmo TAN, algoritmo BAN y el algoritmo HC. Así mismo, la propuesta de la interfaz web, basado en el modelo bayesiano permitió pronosticar la morbilidad del neonato de las madres gestantes, es muy importante tener dicha interfaz web, ya que a la especialista, le permite tomar decisión es acerca de la morbilidad del neonato antes del parto de las madres gestantes.

En esta investigación, la propuesta del Modelo Bayesiano está implementada en función del algoritmo Naive de Bayes, el cual es el adecuado para la predicción de morbilidad del neonato, según las variables de estudio. Este algoritmo está basado en el teorema de probabilidad y utiliza el Teorema de Bayes, la búsqueda y la optimización para una mejor solución ante cualquier problema. Este algoritmo es exhaustivo al momento de hacer las comparaciones y la búsqueda de los padres de variables en estudio, ya que el fin de este algoritmo es la de obtener predicciones de las variables en estudio y así obtener un mejor resultado esperado.

1.2.2 Formulación del problema

A. Problema general

¿De qué manera el aspecto nutricional apoyado por un Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?

B. Problemas específicos.

- Problema específico de la dimensión cereales

¿De qué manera el aspecto nutricional cereales apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?

- Problema específico de la dimensión panes

¿De qué manera el aspecto nutricional Panes apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?

- Problema específico de la dimensión verduras y hortalizas

¿De qué manera el aspecto nutricional verduras y hortalizas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?

- Problema específico de la dimensión frutas.

¿De qué manera el aspecto nutricional frutas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?

- Problema específico de la dimensión aceites y grasas.
¿De qué manera el aspecto nutricional aceites y grasas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?
- Problema específico de la dimensión productos lácteos
¿De qué manera el aspecto nutricional productos lácteos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?
- Problema específico de la dimensión carnes, legumbres y huevos
¿De qué manera el aspecto nutricional carnes, legumbres y huevos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?
- Problema específico de la dimensión complementos vitamínicos
¿De qué manera el aspecto nutricional complementos vitamínicos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?
- Problema específico de la dimensión agua.
¿De qué manera el aspecto nutricional agua apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?
- Problema específico de la dimensión descalcificantes.
¿De qué manera el aspecto nutricional descalcificantes apoyado por el modelo bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?

- Problema específico de la dimensión sal.
¿De qué manera el aspecto nutricional sal apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?
- Problema específico de la dimensión azúcar.
¿De qué manera el aspecto nutricional azúcar apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?
- Problema específico de la dimensión condimentos.
¿De qué manera el aspecto nutricional condimentos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?
- Problema específico de la dimensión tipo de comida.
¿De qué manera el aspecto nutricional Tipo de Comida apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?

1.3. Objetivos:

1.3.1 Objetivo General.

Determinar la manera en que el aspecto nutricional apoyado por un Modelo Bayesiano se relaciona con el Pronóstico de Morbilidad del Neonato de la gestante.

1.3.2 Objetivos específicos.

Determinar la manera en que el aspecto nutricional cereales apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.

- En que forma el aspecto nutricional panes apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.
- De qué forma el aspecto nutricional verduras y hortalizas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.
- En qué forma el aspecto nutricional frutas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.
- De qué manera el aspecto nutricional aceites y grasas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.
- De qué forma el aspecto nutricional productos lácteos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.
- En qué forma el aspecto nutricional carnes, legumbres y huevos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.
- De qué manera el aspecto nutricional complementos vitamínicos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.
- De qué manera el aspecto nutricional agua apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.

- De qué manera el aspecto nutricional descalcificantes apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.
- En que forma el aspecto nutricional Sal apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.
- De qué Manera el aspecto nutricional azúcar apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.
- En que forma el aspecto nutricional condimentos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.
- En que forma el aspecto nutricional tipo de comida apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.

1.4. Hipótesis:

1.4.1 Hipótesis general.

La aplicación del Modelo Bayesiano en el análisis nutricional tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.

1.4.2 Hipótesis específicas.

- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a cereales tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.

- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a panes tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.
- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a verduras y hortalizas tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.
- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a frutas tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.
- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a aceites y grasas tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.
- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a productos lácteos tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.
- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a carnes, legumbres y huevos tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.
- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a complementos vitamínicos tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.
- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a agua tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.

- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a descalcificantes cereales tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.
- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a azúcar tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.
- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto a condimentos tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.
- La aplicación del Modelo Bayesiano en el análisis nutricional en cuanto al tipo comida tiene un impacto positivo en el Pronóstico de Morbilidad del Neonato de la gestante.

1.5 Operacionalización de variables:

1.5.1 Variables dependiente e independiente.

Variable Independiente: Aspecto Nutricional

Variable Interviniente: Modelo Bayesiano

Variable Dependiente: Morbilidad del Neonato

1.5.2 Definición operacional de las variables.

Tabla 1. Operacionalización de variables

Variables	Definición Operacional	Dimensiones	Indicadores	Escalas
Aspecto Nutricional Son alimentos que influyen en la buena y sana nutrición, ya que muchos de los alimentos aportan	Bajo= <400 g/día Normal = 400g/día - 420g /día Alto = >420 g/día	Cereales	1.Trigo 2.Harina de maíz, de avena, harina blanca 3.Arroz integral y blanco 4.Sémola	Bajo<2 Normal=2 Alto>2

<p>proteínas y vitaminas para la salud humana. Cuando se hace un estudio en el aspecto nutricional, se debe de tomar en consideración varios aspectos, dentro de las cuales figura el contexto social, el económico, el político, el histórico, el geográfico y el cultural.</p> <p>Modelo Bayesiano Es una estructura grafica entre n variables independientes y una dependiente, en la que se realizan inferencias a partir de datos, induciendo modelos probabilísticos, el cual analiza numerosas técnicas de aprendizaje y de minería de datos.</p>	<p>Bajo = <400g/día Normal=400g/día – 420g/día Alto = >210g/día</p>	<p>Verduras y hortalizas</p>	<p>1.Lechuga 2.Espinaca 3.Brocoli 4.Zanahorias 5.Calabaza 6.Papas 7.Maíz 8.Garbanzos 9.Lentejas</p>	<p>Bajo=<2 Normal=2 Alto>2</p>
	<p>Bajo<4unidad/día Normal = 4-5unidad/día Alto>5unidad/día</p>	<p>Frutas</p>	<p>1.Plátanos 2.Manzana 3.Naranja 4.Arándanos, frutas secas 5.Pera 6.Kiwi 7.Mangos 8.Uva 9.Melón, piña</p>	<p>Bajo<4 Normal =5 Alto>5</p>
	<p>Bajo <0g/día Normal=0-1g/día Alto> 1g/día</p>	<p>Aceites y grasas</p>	<p>1.Aceite de oliva 2.Aceite de maíz 3.Aceite de girasol 4.Aceite de soja 5.Aceite de almendras 6.Manteca 7.Mantequilla 8.Margarina</p>	<p>Bajo<1 Normal = 1 Alto>1</p>
	<p>Bajo<400g/día Normal=400-420/día Alto>420g/día</p>	<p>Productos lácteos</p>	<p>1.Leches enteras y descremadas 2.Postres y lácteos 3.Quesos 4.Yogures</p>	<p>Bajo<2 Normal=2 Alto>2</p>
	<p>Bajo>200g/día Normal=200-300g/día Alto>300 g/día</p>	<p>Carnes, legumbres y huevos</p>	<p>1.Carne de vaca 2.Carne de cerdo 3.Carne de cordero 4.Mariscos 5.Pavo 6.Pollo 7.Pescados 8.Huevos</p>	<p>Bajo<2 Normal=3 Alto>3</p>

Morbilidad del Neonato Es la muerte producida entre el nacimiento hasta los 28 días de vida. La tasa de mortalidad neonatal es un indicador sensible para valorar la calidad de la atención en salud del recién nacido en una determinada área geográfica o en un servicio y es un indicador de impacto de los programas e intervenciones en esta área.			9.Otros	
	Bajo<1 Normal=1-2unidad/día Alto>2	Complementos vitamínicos	1.Ácido Fólico 2.Otras vitaminas	Bajo<1 Normal=2 Alto>2
	Bajo<7 vasos/día Normal= 7-8vasos/día Alto>8vasos/día	Agua	1.Agua	Bajo<7 Normal=8 Alto>8
	Bajo<0 Normal=0 Alto>0	Descalcificantes	1.Cafeinas 2.Gaseosa 3.Vino 4.Alcohol 5.Alimentos Procesados	Bajo<0 Normal=0 Alto>0
	Desayuno (6:30 - 9:00) Media Mañana (9:30 - 12) Almuerzo (12:30pm - 2:30 pm) Media Tarde (3 pm - 6pm) Cena(6:30pm - 8:00Pm) Noche (9 pm - 12 am)	Tipo de alimentación	1. Desayuno 2.Media Mañana 3.Almuerzo 4.Media Tarde 5.Cena 6.Noche	Desayuno=1 Media Mañana=2 Almuerzo=3 Media Tarde =4 Cena=5 Noche=6
	Bajo<2 Normal= 2-3unidad/día Alto>3	Panes	1.Panes	Bajo<2 Normal=2 Alto>2
	Bajo<0g/día Normal=20-24g/día Alto>24g/día	Azúcar	1.Azúcar	Bajo<6 Normal=6 Alto>6
	Bajo<0g/día Normal =4g/día Alto>4g/día	Sal	1. Sal	Bajo<1 Normal=1 Alto>1
	Bajo< 0g/día Normal = 4g/día Alto>4g/día	Condimentos	1. Condimentos	Bajo<1 Normal=1 Alto>1
	Muy Alto=0.75 - 100 Alto=0.50 -0.74 Medio=0.25 - 0.49 Bajo=0 - 0.24	Morbilidad	1.Morbilidad	Muy Alto=4 Alto=3 Medio=2 Bajo=1

Tabla 2. Tabla de medidas por raciones

Medida asignada(estándar)	210g	1
Unidad	1	1
Ración	100g	1
Cucharadas	4g	1

*Las medidas son según las escalas con las que trabaja el sistema.

1.6 Matriz de Consistencia

Tabla 3. Matriz de consistencia

Título: Aspecto Nutricional apoyado por un Modelo Bayesiano y su relación en el pronóstico de Morbilidad del Neonato de la gestante.						
DEFINICION DEL PROBLEMA	OBJETIVO GENERAL	HIPOTESIS GENERAL	JUSTIFICACION	INDICADORES	VARIABLES	INDICES
<p>Problema general ¿De qué manera el aspecto nutricional apoyado por un Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?</p>	<p>Determinar la manera en que el aspecto nutricional apoyado por un Modelo Bayesiano se relaciona con el Pronóstico de Morbilidad del Neonato de la gestante.</p>	<p>Si se aplica el aspecto nutricional apoyado por un Modelo Bayesiano entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.</p>	<p>Este trabajo de investigación se justifica, porque en su estructura se implementó el algoritmo de Naive Bayes, que apoyado por una interfaz web, se tiene como objetivo pronosticar la morbilidad del neonato, el cual es muy importante para la toma de decisiones.</p>	<p>Pronóstico de la morbilidad materna.</p> <ol style="list-style-type: none"> 1.Trigo 2.Harina de maíz, de avena, blanca 3.Arroz integral y blanco 4.Sémola 5.Pan de maíz 	<p>Modelo Bayesiano</p> <p>Pronostico de Morbilidad del Neonato</p> <p>Aspecto Nutricional</p>	<p>Eficiencia = [0 – 100%]</p> <p>Bajo= <400 g/día Normal = 400g/día - 420g /día Alto = >420 g/día</p>
<p>Problema Especifico ¿De qué manera el aspecto nutricional cereales apoyado por el Modelo</p>	<p>Objetivos específicos: Determinar la manera en que el aspecto nutricional cereales apoyado</p>	<p>Hipótesis específicas: Si se aplica el aspecto nutricional cereales apoyado por un Modelo</p>	<p>En la cual las madres gestantes realizan su control mensualmente, se está beneficiando, porque esta interfaz</p>	<ol style="list-style-type: none"> 1.Lechuga 2.Espinaca 3.Brocoli 4.Zanahorias 5.Calabaza 6.Papas 7.Maíz 8.Garbanzos 9.Lentejas 		<p>Bajo = <400 g/día Normal=400g/día – 420g/día Alto = >210g/día</p> <p>Bajo<4unidad/día</p>

<p>Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?</p> <p>¿De qué manera el aspecto nutricional panes apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?</p> <p>¿De qué manera el aspecto nutricional verduras y hortalizas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del</p>	<p>por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>En que forma el aspecto nutricional panes apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>De qué forma el aspecto nutricional verduras y hortalizas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de</p>	<p>Bayesiano entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>Si se aplica el aspecto nutricional panes apoyado por un Modelo Bayesiano entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>Si se aplica el aspecto nutricional verduras y hortalizas apoyado por un</p>	<p>optimizara el mejor control por parte de las madres gestantes de los hospitales.</p>	<p>1.Plátanos 2.Manzana 3.Naranja 4.Arándanos, frutas secas 5.Pera 6.Kiwi 7.Mangos 8.Uva 9.Melón, piña</p> <p>1.Aceite de oliva 2.Aceite de maíz 3.Aceite de girasol 4.Aceite de soja 5.Aceite de almendras 6.Manteca 7.Mantequilla 8.Margarina</p> <p>1.Leches enteras y descremadas 2.Postres y lácteos 3.Quesos 4.Yogures</p>	<p>Normal = 4-5unidad/día Alto >5unidad/día</p> <p>Bajo <0g/día Normal= 0-1g/día Alto> 1g/día</p> <p>Bajo<400g/día Normal=400-420/día Alto>420g/día</p> <p>Bajo >200g/día</p>
---	--	---	---	--	---

<p>Neonato de la gestante?</p> <p>¿De qué manera el aspecto nutricional frutas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?</p> <p>¿De qué manera el aspecto nutricional aceites y grasas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?</p> <p>¿De qué manera el aspecto</p>	<p>Morbilidad del Neonato de la gestante.</p> <p>En qué forma el aspecto nutricional frutas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>De qué manera el aspecto nutricional aceites y grasas apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.</p>	<p>Modelo Bayesiano entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>Si se aplica el aspecto nutricional frutas apoyado por un Modelo Bayesiano entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>Si se aplica el aspecto nutricional aceites y grasas apoyado por un Modelo Bayesiano</p>		<p>1.Carne de vaca 2.Carne de cerdo 3.Carne de cordero 4.Mariscos 5.Pavo 6.Pollo 7.Pescados 8.Huevos 9.Otros</p> <p>1.Ácido fólico 2.Otras vitaminas</p> <p>1.Cafeínas 2.Gaseosa 3.Cigarro 4.Alcohol 5.Alimentos Procesados</p> <p>1. Desayuno 2.Media Mañana 3.Almuerzo 4.Media Tarde</p>		<p>Normal=200-300 g/día Alto>300 g/día</p> <p>Bajo <1 Normal=1-2unidad/día Alto>2</p> <p>Bajo<0 Normal=0 Alto>0</p> <p>Desayuno (6:30 - 9:00) Media Mañana (9:30 - 12)</p>
---	--	---	--	--	--	---

<p>nutricional productos lácteos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?</p> <p>¿De qué manera el aspecto nutricional carnes, legumbres y huevos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?</p> <p>¿De qué manera el aspecto nutricional complementos vitamínicos apoyado por el</p>	<p>De qué forma el aspecto nutricional productos lácteos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>En qué forma el aspecto nutricional carnes, legumbres y huevos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>De qué manera el aspecto nutricional complementos</p>	<p>entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>Si se aplica el aspecto nutricional productos lácteos apoyado por un Modelo Bayesiano entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>Si se aplica el aspecto nutricional carnes, legumbres y huevos apoyado por un Modelo Bayesiano</p>		<p>5.Cena 6.Noche</p> <p>1.Panes</p> <p>1.Azúcar</p> <p>1.Sal</p> <p>1.Condimentos</p>		<p>Almuerzo (12:30pm - 2:30 pm) Media Tarde (3 pm - 6pm) Cena(6:30pm - 8:00Pm) Noche (9 pm - 12 am)</p> <p>Bajo<2 Normal= 2-3unidad/día Alto>3</p> <p>Bajo<0g/día Normal=20-24g/día Alto >24g/día</p> <p>Bajo<0g/día Normal =4g/día Alto >4g/día</p> <p>Bajo< 0g/día Normal = 4g/día Alto >4g/día</p>
---	---	---	--	--	--	---

Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?	vitamínicos apoyado por el modelo bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.	entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.		1.Morbilidad		Muy Alto=0.75 - 100 Alto=0.50 -0.74 Medio=0.25 -0.49 Bajo=0 - 0.24
¿De qué manera el aspecto nutricional agua apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?	De qué manera el aspecto nutricional agua apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.	Si se aplica el aspecto nutricional complementos vitamínicos apoyado por un Modelo Bayesiano entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.				
¿De qué manera el aspecto nutricional descalcificantes apoyado por el Modelo Bayesiano se relaciona con el pronóstico de	De qué manera el aspecto nutricional descalcificantes apoyado por el Modelo Bayesiano se relaciona con el pronóstico de	Si se aplica el aspecto nutricional agua apoyado por un Modelo Bayesiano				

Morbilidad del Neonato de la gestante?	Morbilidad del Neonato de la gestante.	entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.				
¿De qué manera el aspecto nutricional sal apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?	En que forma el aspecto nutricional sal apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.	Si se aplica el aspecto nutricional descalcificantes apoyado por un Modelo Bayesiano entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.				
¿De qué manera el aspecto nutricional Azúcar apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?	De qué Manera el aspecto nutricional Azúcar apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.	Si se aplica el aspecto nutricional azúcar apoyado por un Modelo Bayesiano entonces se				
¿De qué manera el aspecto		entonces se				

<p>nutricional Condimentos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?</p> <p>¿De qué manera el aspecto nutricional tipo de comida apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante?</p>	<p>En que forma el aspecto nutricional condimentos apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>En que forma el aspecto nutricional tipo de comida apoyado por el Modelo Bayesiano se relaciona con el pronóstico de Morbilidad del Neonato de la gestante.</p>	<p>relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>Si se aplica el aspecto nutricional condimentos apoyado por un Modelo Bayesiano entonces se relaciona positivamente con el pronóstico de Morbilidad del Neonato de la gestante.</p> <p>Si se aplica el aspecto nutricional tipo de comida apoyado por un Modelo Bayesiano entonces se relaciona</p>				
--	--	---	--	--	--	--

		positivamente con el pronóstico de Morbilidad del Neonato de la gestante.				
--	--	---	--	--	--	--

1.7 Justificación:

Este trabajo de investigación se justifica, porque en su estructura se implementó el algoritmo de Naive Bayes, que apoyado por una interfaz web, se tiene como objetivo pronosticar la morbilidad del neonato, que es muy importante para la toma de decisiones, como: el pronóstico de posibles enfermedades, mejorar el proceso de atención y monitoreo. Así mismo, el área Neonatología en la cual las madres gestantes realizan su control mensualmente, se está beneficiando, porque esta interfaz optimizará el mejor control por parte del personal médico de los hospitales.

1.8 Alcances de la investigación.

El alcance de esta investigación está dirigida a todas las madres gestantes que llevan su control en cualquier hospital.

Además en el área de Neonatología, se le hará más eficaz predecir la morbilidad del neonato porque tendrá una interfaz web que le permitirá pronosticar exactamente la morbilidad del neonato.

Los datos obtenidos para la simulación de la interfaz web fueron de dos madres atendidas en el presente año desde el 01 de Septiembre al 14 del mismo mes, del año en curso.

1.9 Limitaciones de la investigación.

- Pocas madres con disposición de ayudar al proceso de análisis de alimentación para su pronóstico de morbilidad neonatal.
- Bajo compromiso ético nutricional de las madres gestantes, puesto que no se cumplían las raciones nutricionales que se debían consumir en el proceso de nutrición para su neonato.

- Limitada disponibilidad de tiempo para las reuniones con los encargados del área de neonatología.
- No todas las madres gestantes involucradas en la muestra de estudio no cuentan con celulares Smartphone para su control y llenado de datos nutricionales para su control neonatal y esto dificulta el proceso de recojo de información para la validación de las hipótesis de la investigación.

Capítulo II:

Marco Teórico - Conceptual

2.1 Estado del arte (antecedentes)

En la investigación los autores explican, el significado y el uso que se le puede dar a un árbol de clasificación, también conocido como árbol de decisión, su dominio está basada en variables predictoras, el algoritmo que se usa para el árbol de decisión es llamado TDIDT (*Top Down Induction Of Decision Trees*), el cual explica cuál es la rama idónea del árbol, previamente se debe tener un criterio ya establecido, para ver que patrones corresponden a una rama del árbol es por eso llamado algoritmo básico.

Según los investigadores existen otro tipo de algoritmo es denominado ID3, el cual nos explica que debe existir información recíproca entre una variable predictora y una variable clase. En este tipo de algoritmo se hace una prueba de independencia a las variables predictoras y las variables clase, la cual servirá para tomar en cuenta que variables se consideraran en el algoritmo.

El algoritmo C4.5, es una mejora de la anterior que utiliza el criterio de ratio ganancia. El objetivo de este algoritmo es evitar que las variables con mayor número de posibles valores salgan beneficiados en la selección, el algoritmo también incluye el criterio de poda del árbol que consiste en hacer un test de prueba, la cual indicará si es necesario expandir o no una determinada rama (Pedro Larrañaga, 2000).

La red bayesiana es una representación gráfica de Nodos Y Arcos que se utilizan para tomar decisiones a través de eventos e incertidumbres de un suceso. Las redes bayesianas utilizan diferentes tipos de algoritmos para diferente tipo de análisis probabilísticos, cada uno de ellos de acuerdo a la necesidad del problema donde el

factor humano muchas veces entra en contradicciones y toma decisiones al azar, sin medir el factor de riesgo en un área determinada de la empresa de cualquier rubro.

Según el trabajo de investigación, describe cómo es que el teorema ayuda a solucionar problemas que muchas veces no tienen solución al ojo humano. Fue creada en el siglo XVII por el conocido Thomas Bayes.

El Teorema de Bayes son modelos gráficos probabilísticos, que se utilizan en la toma de decisiones, para trabajar con variables representativas sobre un conjunto infinito de variables.

Lo que se sabe es que la red bayesiana se pueden comportar de dos tipos: red bayesiana continúa (con valores numéricos infinitos) y las redes bayesianas Dinámicas (contienen variables aleatorias).

Para la aplicación del modelo Bayesiano gráficamente, existen diferentes tipos de herramientas tales como Elvira, entre otros.

El modelado y el adiestramiento de las redes bayesianas, dependiendo de la herramienta que se use, va tener diferente tipos y maneras de trabajar y diferente resultado en cuanto al tipo de algoritmo que se le implemente a la herramienta para el adiestramiento de las redes bayesianas (Gabriela Rodrigues, 2014).

Para modelar una red bayesiana, utilizando una herramienta, podemos utilizar una infinidad de herramientas que existen en el mercado. Una de ellas es Netica, con la cual el presente trabajo hizo el análisis de datos y el modelado de la herramienta.

Las redes bayesianas se han popularizado en la Psicología debido a que es usada para modelar procesos cognitivos.

El trabajo con la red bayesiana se basa en un proceso de construcción, como cuando se generan modelos ecuaciones estructurales.

Existen dos procedimientos genéricos para crear redes bayesianas: uno basado en rutinas automáticas y la otra que se basa en el juicio de personas expertas en el tema.

Para la estimación estructural con el uso del Netica, se trabajó los procedimientos de forma manual, ya que el Netica no incorpora ningún procedimiento automático, para su modelado, para ello se diseñó la estructura grafica de forma manual bajo criterios de estudios previos de redes bayesiana.

Netica para la parametrización utiliza algoritmos que ayudan a estimar una red bayesiana. En este caso el algoritmo que se usa es el algoritmo EM y el algoritmo que evalúa la reducción del gradiente aprendizaje.

Para su respectiva validación, netica evalúa el grado de comportamiento que se ajusta a un conjunto de datos, para una validez cruzada (Jorge López, 2012).

La importancia que tienen las redes bayesianas en la toma de decisiones son criterios que toma en consideración varios aspectos, entre las cuales está el análisis de hechos concretos que se basa en gran medida de la capacidad de hacer inferencias sobre la ocurrencia de eventos futuros.

En el presente artículo, el autor explica acerca de cómo se ejecutan las redes bayesianas, este tipo de redes trabaja utilizando el teorema de Bayes.

Entonces se hace un previo estudio dentro de un esquema financiero para que ayude a controlar y pronosticar los diferentes factores dentro del marketing.

Previamente, el autor hace una pequeña introducción acerca del concepto de una red bayesiana la define como una herramienta informática que puede crear y modelar diferentes esquemas de modelado, dependiendo del caso y el problema que se presente, a fin de observar el comportamiento de las variables.

Muchas veces, dentro del modelado de las redes bayesianas, se habla acerca de la toma de decisiones, puesto que muchas veces dentro de un determinado estudio y análisis existe incertidumbre.

Con la implementación de las redes bayesianas, no solo ayuda a tomar buenas decisiones, sino que muchas veces ayuda a la optimización de un determinado problema.

En la actualidad, las redes bayesianas son implementadas y usadas dentro del rubro bancario, que ayuda no solo a predecir, sino también a tomar decisiones en cuanto al riesgo crediticio (Miller Rivera, 2011).

Según el artículo, lo que se pretende implementar en una escuela es un programa que ayude al problema del aprendizaje mediante imágenes que los niños plasman en un papel mediante pruebas realizadas en el colegio.

Para esto los investigadores se valieron de las herramientas como WEKA, para el procesamiento de la red neuronal, y también se empleó los coeficientes en series de Fourier para el entrenamiento de la red neuronal.

El objeto en estudio fue con respecto a las imágenes que hacían los niños y con esto determinar si el niño tiene o padece de algún tipo de problema en cuanto a aprendizaje, para ello se tuvo que entrenar la red neuronal con el fin de que pudiera clasificar entre las imágenes diseñadas por los niños (entre 3 y 6 años) y determinar si lo que ellos hacen es un círculo o no.

Se tomó como muestra de estudio a niños de tres años entre una población de 43 niños; niños de 4 años entre una población de 31 y, por último, niños de 5 años, entre una población de 19.

Según el artículo, se tomó herramientas para ayudar a generar círculos irregulares de forma aleatoria.

Para este estudio se determinó varios aspectos en las cuales cabe resaltar el tipo de papel a usar para digitalización, y el tipo de lápiz para el trazado de los dibujos, y por último, el uso de un escáner para la digitalización de los dibujos hechos a papel. Todo esto para la alimentación de los datos en la herramienta y su posterior entrenamiento de las redes.

El procesamiento de los datos (imágenes) alimentó la red neuronal para su entrenamiento y posteriormente utilizarla en la clasificación de las figuras, luego de la alimentación se implementó el algoritmo que se usó para su desarrollo.

Una vez entrenado la red neuronal e implementada el algoritmo, se usó la herramienta WEKA para probar y entrenar la red neuronal

Se pudo concluir que hubo algunas dificultades con respecto al tamaño de la población con la que se trabajó (Gabriela Esperón, 2008).

2.2 Marco teórico.

Modelo Bayesiano.

La teoría de la probabilidad y los métodos bayesianos son una de las técnicas que más se han utilizado en problemas de inteligencia artificial y, por tanto, de aprendizaje automático y minería de datos. Como se indica en (Mitchell, 1977) dos son las razones por la que los métodos bayesianos son relevantes al aprendizaje automático y a la minería de datos:

1. Son un método práctico para realizar inferencias a partir de datos, induciendo modelos probabilísticos que después serán usados para razonar (formular hipótesis)

sobre nuevos valores observados. Además, permiten calcular de forma explícita la probabilidad asociada a cada una de las hipótesis posibles, lo que constituye una gran ventaja sobre otras técnicas.

2. Facilitan un marco de trabajo útil para la comprensión y análisis de numerosas técnicas de aprendizaje y minería de datos que no trabajan explícitamente con probabilidades.

Teorema de Bayes e hipótesis MAP.

En teoría de probabilidad, el teorema de Bayes es la regla básica para realizar inferencias. Así, el teorema de Bayes nos permite actualizar a creencia que tenemos en un suceso o conjunto de sucesos a la luz de nuevos datos u observaciones. Es decir, nos permite pasar de la probabilidad a priori P (suceso) a la probabilidad a posteriori p (suceso/observaciones). La probabilidad a priori puede verse como la probabilidad inicial, la que fijamos sin saber nada más.

Teniendo en cuenta estos conceptos, el teorema de Bayes viene representado por la siguiente expresión:

$$P(h/O) = \frac{P(O/h).P(h)}{P(O)}$$

Como podemos ver, lo que aparecen son la probabilidad a priori de la hipótesis (h) y de las observaciones (O) y de las probabilidades condicionadas $P(h/O)$ y $P(O/h)$. A esta última se le conoce como la verosimilitud de que la hipótesis h haya producido el conjunto de observaciones O .

Centrándonos en el problema de la clasificación, con una variable clase (C) y un conjunto de variables predictoras o atributos $\{A_1, \dots, A_n\}$ el teorema de Bayes tendría la siguiente forma:

$$P(C/A_1 \dots A_n) = \frac{P(A_1 \dots A_n / C) \cdot P(C)}{P(A_1 \dots A_n)}$$

Evidentemente, si (C) tiene k posibles valores $\{C_1, \dots, C_k\}$, lo que nos interesa es identificar el más plausible y devolverlo como resultado de la clasificación. En el marco bayesiano, la hipótesis más plausible no es otra que aquella que tiene máxima probabilidad a posteriori dados los atributos, y es conocida como la hipótesis máxima a posteriori o hipótesis MAP. Así la clase o valor a devolver será:

$$C_{MAP} = \arg_{c \in \Omega_c} \max P(C/A_1 \dots A_n) = \arg_{c \in \Omega_c} \max \frac{P(A_1 \dots A_n / c) P(c)}{p(A_1 \dots A_n)}$$

$$p(A_1 \dots A_n / c) P(c)$$

Redes Bayesianas.

Las redes bayesianas, en los últimos años ha demostrado su potencialidad como modelo de representación del conocimiento con incertidumbre. Este formalismo nació como una aportación de diferentes campos de investigación: teoría de toma de decisiones, estadística e inteligencia artificial. El éxito de numerosas aplicaciones en campos variados como la medicina, la recuperación de información, la visión artificial, la fusión de información, la agricultura, etc., la avalan este formalismo.

Las redes bayesianas representan el conocimiento cualitativo del modelo mediante un grafo dirigido acíclico. Este conocimiento se articula en la definición de

relaciones de independencia/dependencia entre las variables que componen el modelo. Estas relaciones abarcan desde una independencia completa hasta una dependencia funcional entre variables del modelo. El hecho de utilizar una representación gráfica para la especificación del modelo hace de las redes bayesianas una herramienta realmente muy atractiva en su uso como representación del conocimiento, aspecto muy importante de la minería de datos.

Las redes bayesianas no solo modelan de forma cualitativa el conocimiento sino que además expresan de forma numérica la “fuerza” de las relaciones entre las variables. Esta parte cuantitativa del modelo suele especificarse mediante distribuciones de probabilidad como una medida de la carencia que tenemos sobre las relaciones entre variables del modelo.

Formalmente, una red bayesiana es una tupla $B = (G, \Theta)$, donde G es el grafo y Θ es el conjunto de distribuciones de probabilidad $P(X_i/Pa(X_i))$ para cada variable desde $i = 1$ hasta n y $Pa(X_i)$ representan los padres de la variable X_i en el grafo G .

Veremos un ejemplo de una red bayesiana:

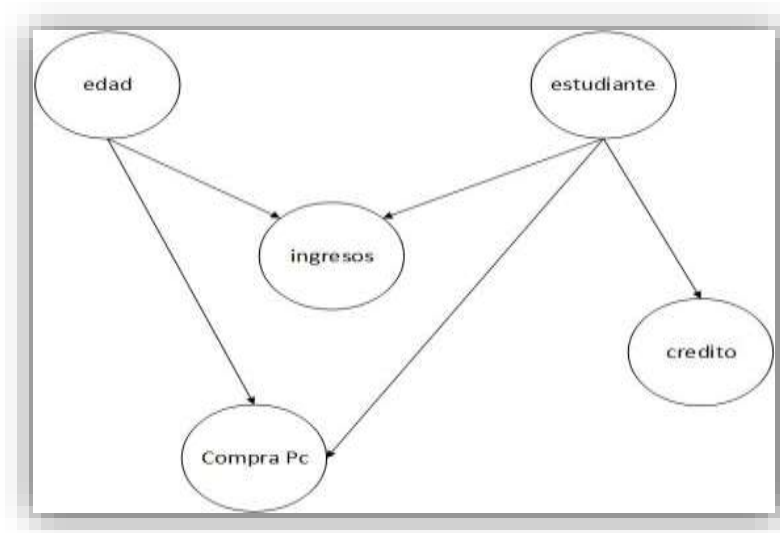


Figura 1. Red bayesiana compra de un equipo informático

En el ejemplo se puede observar que la presencia de un arco directo entre dos variables expresa una relación directa entre las variables que une; si se interpreta esta relación, es una relación causa – efecto, podremos expresar de forma explícita una manera de representar conocimiento muy habitual entre los expertos en sus determinados dominios. Pero además, mediante una red bayesiana podremos extraer relaciones de independencia no directas entre las variables con el criterio conocido como d-separación o separación dirigida.

Podemos observar que el nivel de ingresos depende directamente de la edad y de su condición de estudiante, la compra de una computadora personal depende directamente de la edad, la condición del estudiante y el nivel de crédito. Otra relación no directa que describe nuestro ejemplo es que la edad es (marginamente) independiente del nivel de crédito y que los ingresos son independientes de comprar una computadora personal conocida la edad.

El formalismo que se va utilizar para representar la parte cuantitativa del modelo red, es la teoría de la probabilidad. En entornos probabilísticos, una distribución de probabilidad P puede ser considerada un modelo de dependencias utilizando la siguiente relación:

$$I(X,Y/Z) \Leftrightarrow P(X/YZ) = P(X/Z)$$

Donde X, Y, Z son subconjuntos de variables del modelo y la sentencia $I(.,.I.)$ se interpreta como una relación de independencia condicional. La expresión anterior se leería como “ X es condicionalmente independiente de Y conocido Z ”.

En un entorno probabilístico, el conocimiento cuantitativo viene expresado por una distribución de probabilidad conjunta definida sobre las variables del modelo. La codificación de las relaciones de independencia condicional expresadas en una red bayesiana hace que la distribución de probabilidad conjunta se pueda almacenar de una manera mucho más eficiente y local a cada una de las variables del modelo, esto es:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i / Pa(X_i)).$$

Si nosotros tuviésemos n variables aleatorias binarias (discretas con dos valores), necesitaríamos almacenar $O(2^n)$ parámetros para especificar la distribución de probabilidad conjunta $P(X_1, X_2, \dots, X_n)$. Sin embargo, una red bayesiana puede necesitar un número exponencialmente menor de parámetros, dado que para cada variable tan solo hay que almacenar una distribución de probabilidad condicionada a su conjunto de padres en el grafo previamente especificado.

Una vez que hemos definido las redes bayesianas, se pueden destacar dos problemas fundamentales para trabajar con este formalismo. Por un lado, se trata de

utilizar estas para realizar procesos eficientes de razonamiento una vez que tengamos especificado el modelo completo tanto cuantitativa como cuantitativamente. Para este tipo de tareas han sido muchos los trabajos realizados [Jensen 2001] y veremos a continuación un ejemplo. La existencia de estos algoritmos cada vez más eficientes hace que aparezcan cada vez más aplicaciones prácticas muy interesantes que utilizan redes bayesianas como motor de inferencia en sus sistemas. Por otro lado, el segundo problema es la tarea de su construcción.

Aprendizaje de redes bayesianas.

Una red bayesiana, una vez construida, constituye un dispositivo potente para el razonamiento probabilístico. Sin embargo, nos queda la tarea de la construcción de tal modelo. Sin embargo, nos queda la tareade la construcción de tal modelo. Una posibilidad es que un experto, en el dominio que se quiere modelar, construya la red bayesiana a partir de su conocimiento en el problema. Debido al gran volumen de datos de los que habitualmente se dispone en dominios concretos, es de enorme interés proporcionarles a estos expertos herramientas que adquieran este tipo de conocimiento de forma automática a partir de datos de ejemplos del problema en cuestión, para que de esta forma tengan herramienta de soporte para la decisión.

El problema del aprendizaje de redes bayesianas se puede definir como sigue: dado un conjunto de datos D , encontrar el grafo dirigido acíclico G que mejor represente el conjunto de dependencias/independencias presentes en los datos. Para solucionar este problema, desde el punto de vista bayesiano, es necesario calcular la probabilidad a posteriori de una red bayesiana en concreto, dado el conjunto de datos conocidos, esto es, $P(G/D)$. Una vez que sabemos cómo calcular esta probabilidad,

tendremos una medida de adecuación de cada grafo (red bayesiana) a los datos de partida y por consiguiente podremos comparar, para quedarnos con la mejor, entre distintas redes bayesianas.

Por otra parte, es conocido que el conjunto de redes bayesianas con n nodos es de orden súper-exponencial, con lo que la enumeración exhaustiva del conjunto de redes para encontrar la mejor red candidata no es factible, además se ha demostrado que el problema del aprendizaje, tal como lo hemos definido, es un problema NP-duro. Como acabamos de comentar, no es posible, por tanto, la resolución exacta a nuestro problema, por consiguiente, es habitual utilizar algún mecanismo de búsqueda heurística para guiar el problema y así encontrar cada vez mejores redes bayesianas. Para resolver el problema, en primer lugar vamos a definir algunas medidas o métricas bayesianas de la optimización de una red bayesiana respecto a unos datos.

Medidas bayesianas.

Sea $G=(V,E)$ un grafo dirigido acíclico que representa una red bayesiana $B=(G,\Theta)$, donde Θ es el conjunto de parámetros definidos en la red, esto es, el conjunto de distribuciones de probabilidad condicional. Sean D los N casos de datos con la siguiente forma:

$$D = \begin{bmatrix} X_1[1] & X_2[1] & \dots & X_n[1] \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ X_1[N] & X_2[N] & \dots & X_n[N] \end{bmatrix}$$

El problema consiste en calcular la probabilidad $B=(B/D)$, es decir la probabilidad de una red dados los datos. Aplicando la regla de Bayes, obtenemos que:

$$P(G/D) = \frac{P(D/G)P(G)}{P(D)}$$

Como los datos son conocidos y constantes, entonces el término $P(D)$ puede eliminarse de la expresión anterior. Además, una red bayesiana posee dos componentes (parámetros y estructura o grafo) con lo que la expresión anterior se traduce en:

$$P(G/D) \propto P(G)L(D/G)$$

$$L(D/G) = \int_{\Theta} P(D/G, \Theta)P(\Theta/G)d\Theta$$

Siendo $L(D/G)$ la función denominada verosimilitud marginal. Para el caso discreto, asumiendo que las distribuciones son de la familia exponencial y en concreto para el caso multinomial y su conjugada, la formula anterior tiene una solución cerrada dando lugar a una medida bayesiana, conocida como BDe:

$$L(D/G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}$$

Siendo Γ la función gamma y α_{ijk} los hiper parámetros de la distribución a priori. Si estos hiper parámetros se suponen uniformes, entonces los α_{ijk} pueden igualarse a una constante, denominada tamaño muestral equivalente. Además, N_{ijk} es la frecuencia de aparición en los datos de la variable i -ésima con su k -ésimo valor (de los r_i posibles) y para la configuración j -ésima (de los q_i posibles) de sus padres en el grafo. Por último, tanto N_{ijk} como α_{ijk} es igual a su proyección sobre k , esto es:

$N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. Para un estudio más profundo sobre las suposiciones realizadas y los pasos hasta la expresión anterior.

Hemos de destacar que la métrica es descomponible, en el sentido de que se puede calcular de forma local para cada familia (conjunto formado por un nodo y sus padres en G). Por tanto, la métrica podrá expresarse, tomando la función logarítmica, como:

$$f(G : D) = \log L(D/G) = \sum_{i=1}^n f(X_i / Pa(X_i) : D)$$

Siendo:

$$f_i(X_i / Pa(X_i) : D) = \sum_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} + \sum_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})}$$

Existen otros tipos de medidas de calidad para calcular la adecuación de una red bayesiana a un conjunto de datos. Estas se basan en otros enfoques provenientes de la teoría de la información y del principio de longitud mínima, en donde se utilizan información y del principio de longitud de descripción mínima, en donde se utilizan criterios de máxima verosimilitud para la estimación de los parámetros en una red bayesiana, como por ejemplo la métrica denominada BIC (Bayesian Information Criterion), la cual tiene la siguiente expresión:

$$f_i(X_i / Pa(X_i) : D) = \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{ijk} \log \frac{N_{ijk}}{N_{ij}} - \frac{1}{2} Dim(B) \log N_{ss}$$

Donde $Dim(B)$ representa una función para calcular el número de parámetros libres que se necesitan especificar en una red bayesiana, esto es, $Dim(B) = \sum_{i=1}^n (r_i - 1) q_i$ (número de entradas independientes en las tablas de probabilidad) y N representa el número total de casos en los datos.

Morbilidad del Neonato.

Según [Asamblea Mundial de la Salud, resolución WHA20.19 y WHA43.24, 2007] El período neonatal comienza en el nacimiento y termina 28 días completos después del nacimiento. Las muertes neonatales (las muertes entre los nacidos vivos durante los primeros 28 días completos de vida) pueden subdividirse en muertes neonatales precoces, que ocurren durante los siete primeros días de vida, y muertes neonatales tardías, que ocurren después del séptimo día pero antes de los 28 días completos de vida Pronóstico.

2.3 Marco conceptual.

En el presente marco conceptual, se presentan conceptos, en relación al aspecto nutricional, a los modelos bayesianos y a los pronósticos de morbilidad materna

- **Cereales.**

Son gramíneos, herbáceos cuyos granos o semillas están en la base de la alimentación. Los cereales contienen almidón y carbohidratos, que es el componente principal de los alimentos humanos.

Verduras y hortalizas.

Estos alimentos constituyen, junto con las frutas, una fuente primordial de vitaminas variadas, sales minerales, fibra y elementos antioxidantes. El bajo consumo de estos alimentos tiene relación con la aparición y desarrollo de algunas de las enfermedades crónicas de mayor prevalencia en la actualidad: aterosclerosis, diverticulitis, colon irritable, osteoporosis, anemias, cáncer, etc.

Las vitaminas son sustancias que en su mayor parte el organismo no puede sintetizar y, aunque necesarias en muy pequeña cantidad, su ingesta inadecuada puede

producir alteraciones funcionales, orgánicas y clínicas. En la actualidad, se destaca su influencia en el freno del desarrollo de la mayor parte de las enfermedades crónicas.

- **Frutas.**

Son frutos comestibles obtenidos de plantas cultivadas o silvestres que, por su sabor generalmente dulce-acidulado, por su aroma intenso y agradable, y por sus propiedades nutritivas, suelen consumirse mayormente en su estado fresco, como jugo o como postre (y en menor medida, en otras preparaciones), una vez alcanzada la madurez organoléptica, o luego de ser sometidos a cocción.

El consumo de frutas aporta pocas calorías y un alto porcentaje de agua (entre 80 y 95 % de su peso fresco), por lo que facilita la hidratación del organismo. Coadyuva al correcto funcionamiento del aparato digestivo por el aporte de fibra alimentaria. Salvo excepciones (por ejemplo, el coco y el aceite de palma obtenido del fruto de *Elaeis guineensis*), las frutas no aportan grasas saturadas. Algunas frutas son fuentes de ácidos grasos esenciales para el organismo, tales como los frutos secos y las paltas o aguacates. Las frutas son además una importante fuente de energía para el organismo por su alto contenido en hidratos de carbono solubles de rápida disponibilidad.

Como alimento, las frutas realizan aportes a la dieta que son de suma importancia para la salud humana. En general, son ricas en vitaminas, minerales y antioxidantes. Algunas vitaminas y minerales, como la vitamina C y el potasio, dependen en buena medida de las frutas como fuentes de suministro. Existen frutas como la sandía o el melón que contienen un alto índice de agua. Además, la sandía aporta al organismo hasta un 20 % del agua que necesita.

- **Aceites y grasas.**

Son triglicéridos o también llamados ésteres de la glicerina, con ácidos grasos de larga cadena de hidrocarburos que generalmente varían en longitud. Aportan calorías (kcal), ácidos grasos esenciales, como el Omega 3, y transportan vitaminas A, D, E y K. Es necesario consumir grasas y aceites de buena calidad, todos los días, en cantidades moderadas.

Estos se dividen en:

Ácidos grasos saturados. Están presentes en alimentos de origen animal, como carnes rojas (vacuno, cerdo, cordero), lácteos (leche entera y quesos), cremas, mantequillas, mayonesas, y alimentos procesados como paté, cecinas, papas fritas, galletas y pasteles.

Ácidos Grasos Trans. Son grasas que se forman al procesar aceites vegetales mediante hidrogenación para aumentar su vida útil y modificar su consistencia de líquida a sólida. Se encuentra en pequeñas cantidades, en la leche y la grasa corporal de los rumiantes como el vacuno.

- **Productos lácteos.**

Es un alimento, deriva de la leche (como el queso o el yogur).

La leche y los alimentos derivados (que por lo general se obtienen a partir del procesamiento y la fermentación de la leche) son altamente perecederos y deben conservarse refrigerados. Por eso, es importante que se mantenga la cadena de frío hasta que el producto llega al consumidor final.

Muchos son los tipos de lácteos que existen en el mercado. No obstante, para poder clasificarlos se toma como punto de partida la leche cruda y, a partir de ella, se establecen cuatro grandes grupos:

- Leche en polvo.
 - Leche desnatada, dentro de la cual hay caseínas y leche desnatada en polvo.
 - Crema de leche. En este apartado, se encuentran tanto las grasas lácteas como la mantequilla.
 - Queso. En este caso, en el grupo se establecen además las caseínas y el suero de leche.
- Y es que gracias a la unión de ambas se consigue elaborar el citado producto, que se puede adquirir en el mercado en una gran variedad de propuestas: fresco, curado, semicurado y añejo.

En la actualidad, la mayor parte de los lácteos proceden de la vaca. De todas maneras, también se consume leche de oveja, cabra, búfala y otros animales. Debido a sus propiedades nutricionales, los lácteos son alimentos populares en todo el mundo. La leche, la manteca o mantequilla y la crema de leche o nata son lácteos sin fermentación, que se obtienen a partir de la adición de nutrientes, de la separación de sus contenidos grasos o de otros procesos. Entre los lácteos con fermentación, se destacan el yogur y el queso.

Cabe destacar que la leche contiene lactosa, un disacárido que no es tolerado por quienes tienen deficiencias en la producción de una enzima conocida como lactasa. Por eso se dice que estas personas sufren de intolerancia a la lactosa.

Legumbres.

Son semillas que están contenidas en las plantas de la familia de las Leguminosas (Fabaceae).

Las legumbres constituyen un grupo de alimentos muy homogéneo, desarrollados a partir del gineceo, de un solo carpelo y que se abre tanto por la sutura ventral como por el nervio dorsal, en dos valvas y con las semillas en una hilera ventral.

Estas vainas suelen ser rectas y carnosas. Por lo general poseen una carne interior esponjosa, aterciopelada y de color blanco. Su parte interna corresponde al mesocarpio y al endocarpio del fruto.

El tamaño de las legumbres varía desde un milímetro o poco más hasta cincuenta centímetros. Su forma, aunque en la mayoría de los casos es alargada y comprimida, como la de las judías, frijoles o habichuelas, varía muchísimo.

La parte de la planta consumida en alimentación animal y humana varía entre las distintas especies de leguminosas. En la mayor parte de los casos, la parte comestible coincide con la utilizada por la planta como almacén de sustancias de reserva. La gran variación existente en la parte consumida es una consecuencia de la diversidad de estrategias utilizadas por las leguminosas para su adaptación a los medios más diversos.

Entre las principales legumbres recomendadas para la alimentación, tenemos las siguientes:

Alverja (*Pisum sativum*)

Fríjol, poroto, alubia o habichuela (*Phaseolus vulgaris*)

Garbanzos (*Cicer arietinum*)

Habas (*Vicia faba*)

Vainita o poroto verde (*Phaseolus vulgaris*)

Lentejas (*Lens culinaris*)

Chocho (*Lupinus*)

Maní (*Arachis hypogaea*)

Soja o soya (*Glycine max*)

Complemento vitamínico.

Son vitaminas que ayudan a la gestante durante su proceso de gestación, a fin de completar las vitaminas que necesita su organismo y suplir la carencia de algunos componentes que el cuerpo necesita durante todo el proceso de gestación.

Desde el comienzo del embarazo, el médico recomienda a las madres gestantes, ácido fólico y yodo.

El ácido fólico, de hecho, se recomienda empezar a tomar unos 3 meses antes del embarazo, para ayudar a prevenir en el desarrollo del bebé los defectos de tubo neuronal (defectos congénitos del cerebro y la médula espinal como la espina bífida y la anencefalia), defectos cardíacos congénitos y malformaciones del aparato urinario. Nuestro cuerpo no produce por sí solo el ácido fólico necesario para llegar a los 400 mcg/día y por ello debemos complementar con un suplemento específico.

El yodo en el ser humano es fundamental para el correcto funcionamiento de la hormona tiroidea y prevenir el hipotiroidismo. Durante el embarazo aumentan los requerimientos hormonales dados la naturaleza del propio estado y debemos suplementarlos de manera externa.

Mantener los niveles adecuados de la hormona tiroidea es la que ayudará al crecimiento y maduración del sistema nervioso central del bebé.

Agua.

Es una sustancia cuya molécula está formada por dos átomos de hidrógeno y uno de oxígeno (H₂O). Es esencial para la supervivencia de todas las formas conocidas de vida. Al nacer, aproximadamente el 75% de nuestro cuerpo es agua y en la edad adulta este porcentaje disminuye hasta aproximadamente el 60%.

Cumple numerosas funciones en nuestro cuerpo, desde la regulación de la temperatura hasta su intervención en numerosos procesos metabólicos.

Podemos distinguir distintas funciones del agua, entre las que destacan las siguientes:

- Se encarga de transportar nutrientes al interior de las células, ayudando a la vez a eliminar los productos de desperdicio de éstas.
- Participa en la digestión de nutrientes.
- Ayuda a la disolución de los mismos junto con la digestión mediante la hidrólisis.
- Interviene en dar volumen y forma a las células. Porque el agua en el interior de las mismas ejerce una presión dándoles la forma adecuada.
- Evita el roce en las articulaciones.
- Mantiene la temperatura corporal mediante el sudor.
- Es el medio líquido donde se encuentran disueltos todos los fluidos corporales como la sangre, orina, heces, secreciones digestivas, etcétera.

Ácido Fólico.

El ácido fólico, folacina o ácido pteroil-L-glutámico (la forma aniónica se llama folato), conocida también como vitamina B9-2 es una vitamina hidrosoluble del complejo de vitaminas B, necesaria para la formación de proteínas estructurales y hemoglobina (y por esto, transitivamente, de los glóbulos rojos); su insuficiencia en los humanos es muy rara.

Las causas de su carencia son la mala alimentación y un déficit de hidratación del folato genético que es asintomático hasta que la mujer se queda embarazada.

Si la mujer tiene suficiente ácido fólico en el cuerpo antes de quedarse embarazada, esta vitamina puede prevenir deformaciones en la placenta que supondrían

el aborto, defectos de nacimiento en el cerebro (anencefalia) y la columna vertebral (espina bífida) del bebé por mal cierre del tubo neural en los extremos cefálico y caudal respectivamente. La espina bífida, un defecto de nacimiento en la columna, puede producir la parálisis de la parte inferior del cuerpo, la falta de control del intestino y la vejiga, y dificultades en el aprendizaje. Si el feto sufre déficit de ácido fólico durante la gestación también puede padecer anemia megaloblástica, ser prematuro o presentar bajo peso al nacer. La madre puede sufrir eclampsia, un proceso que cursa con hipertensión y albuminuria. El ácido fólico también ayuda a mantener un útero sano.

Vitaminas.

Son compuestos heterogéneos imprescindibles para la vida, que al ingerirlos de forma equilibrada y en dosis esenciales promueven el correcto funcionamiento fisiológico. La mayoría de las vitaminas esenciales no pueden ser sintetizadas (elaboradas) por el organismo, por lo que éste no puede obtenerlas más que a través de la ingesta equilibrada de vitaminas contenidas en los alimentos naturales.

Las vitaminas son nutrientes que junto con otros elementos nutricionales actúan como catalizadoras de todos los procesos fisiológicos (directa e indirectamente).

Las frutas y verduras son fuentes importantes de vitaminas.

Los requisitos mínimos diarios de las vitaminas no son muy altos, se necesitan tan solo dosis de miligramos o microgramos contenidas en grandes cantidades (proporcionalmente hablando) de alimentos naturales. Tanto la deficiencia como el exceso de los niveles vitamínicos corporales pueden producir enfermedades que van desde leves a graves e incluso muy graves como la pelagra o la demencia entre otras, e incluso la muerte.

La deficiencia de vitaminas se denomina avitaminosis mientras que el nivel excesivo de vitaminas se denomina hipervitaminosis.

Está demostrado que las vitaminas del grupo B son imprescindibles para el correcto funcionamiento del cerebro y el metabolismo corporal. Este grupo es hidrosoluble (solubles en agua) debido a esto son eliminadas principalmente por la orina, lo cual hace que sea necesaria la ingesta diaria y constante de todas las vitaminas del complejo "B" (contenidas en los alimentos naturales).

Las vitaminas se pueden clasificar según su solubilidad: si lo son en agua hidrosolubles o si lo son en lípidos liposolubles. En los seres humanos hay 13 vitaminas que se clasifican en dos grupos: (9) hidrosolubles (8 del complejo B y la vitamina C) y (4) liposolubles (A, D, E y K).

Vitaminas liposolubles. Las vitaminas liposolubles, A, D, E y K, se consumen junto con alimentos que contienen grasa.

Son las que se disuelven en grasas y aceites. Se almacenan en el hígado y en los tejidos grasos, debido a que se pueden almacenar en la grasa del cuerpo no es necesario tomarlas todos los días por lo que es posible, tras un consumo suficiente, subsistir una época sin su aporte.

Si se consumen en exceso (más de 10 veces las cantidades recomendadas) pueden resultar tóxicas. Esto les puede ocurrir sobre todo a deportistas, que aunque mantienen una dieta equilibrada recurren a suplementos vitamínicos en dosis elevadas, con la idea de que así pueden aumentar su rendimiento físico. Esto es totalmente falso, así como la creencia de que los niños van a crecer más si toman más vitaminas de las necesarias.

Las vitaminas liposolubles son:

Vitamina A (retinolftalina)

Vitamina D (calciferol)

Vitamina E (tocoferol)

Vitamina K (antihemorrágica)

Estas vitaminas no contienen nitrógeno, son solubles en grasa, y por tanto, son transportadas en la grasa de los alimentos que la contienen.

Vitaminas hidrosolubles

Son aquellas que se disuelven en agua. Se trata de coenzimas o precursores de coenzimas, necesarias para muchas reacciones químicas del metabolismo.

En este grupo de vitaminas, se incluyen las vitaminas B1 (tiamina), B2 (riboflavina), B3 (niacina o ácido nicotínico), B5 (ácido pantoténico), B6 (piridoxina), B7/B8 (biotina), B9 (ácido fólico), B12 (cobalamina) y vitamina C (ácido ascórbico).

Estas vitaminas contienen nitrógeno en su molécula (excepto la vitamina C) y no se almacenan en el organismo, a excepción de la vitamina B12, que lo hace de modo importante en el hígado. El exceso de vitaminas ingeridas se excreta en la orina, por lo cual se requiere una ingesta prácticamente diaria, ya que al no almacenarse se depende de la dieta. Por otro lado, estas vitaminas se disuelven en el agua de cocción de los alimentos con facilidad, por lo que resulta conveniente aprovechar esa agua para preparar caldos o sopas.

Descalsificantes.

Son alimentos que absorben el calcio o que lo eliminan de nuestro organismo:

Proteínas animales: en una dieta donde predomina la carne roja, las aves y los huevos suele haber acidosis metabólica; es decir, un desarreglo en el cuerpo que causa una mayor eliminación de calcio en la sangre.

Sodio: comer con mucha sal aumenta la excreción de calcio mediante la orina. Para evitar ello, es necesario dejar de consumir productos enlatados, snacks salados, comidas ya preparadas y empanadas, fast food. Es importante reducir el consumo al preparar los alimentos y en lo posible no llevar el salero a la mesa. La cantidad adecuada por día es de hasta 2 gramos.

Tabaco: si bien no es un alimento, también está dentro de los descalcificantes más importantes. Las personas que fuman son más vulnerables a perder calcio, sobre todo en el caso de las mujeres mayores de 40 años; es decir, después de la menopausia.

Refrescos azucarados: las bebidas cola son muy malas para el cuerpo por diferentes razones. Además de tener mucho azúcar y efervescencia, tienen un alto contenido de fósforo (en forma de ácido fosfórico). Este mineral en poca cantidad es bueno, pero en los refrescos supone efectos contrarios. Favorece, igual que la carne, a la acidosis.

Alcohol, café, productos refinados (pan, arroz, harina y azúcar blancos), porque favorecen a la eliminación de minerales como el calcio del organismo.

IMC.

Es una medida de asociación entre la masa y la talla de un individuo ideada por el estadístico belga Adolphe Quetelet, por lo que también se conoce como índice de Quetelet.

Se calcula según la expresión matemática:

$$IMC = \frac{masa}{estatura^2}$$

La unidad de medida del IMC en el sistema MKS:

$$\text{Kg} * m^2 = \frac{\text{Kg.}}{m^2}$$

El valor obtenido no es constante, sino que varía con la edad y el sexo. También depende de otros factores, como las proporciones de tejidos muscular y adiposo. En el caso de los adultos se ha utilizado como uno de los recursos para evaluar su estado nutricional, de acuerdo con los valores propuestos por la Organización Mundial de la Salud, según el cuadro de comparación:

Horario de alimentación.

Es el horario que las personas le dedican a la alimentación para el buen funcionamiento del organismo. Según los profesionales, en la actualidad se recomienda tener al menos tres comidas estables y dos colaciones. Eso es lo que hoy se entiende como una alimentación equilibrada.

Hoy en cambio, la rapidez con la que nos movemos impide muchas veces que algunos se tomen el tiempo para alimentarse bien y con esto, se saltan las comidas, picotean lo primero que encuentran y recurren a la no muy bien ponderada fast food, comida que se convirtió ya en un clásico en gran parte de la población.

De esta manera, la carencia de nutrientes, el sobrepeso y otras enfermedades asociadas con una alimentación desequilibrada – incluso el mal rendimiento intelectual – pueden tener directa relación con haber olvidado que existe la concepción de las cuatro comidas, todas con sus propias cualidades y razones de existir.

- **Panes.**

El pan forma parte del grupo de alimentos que han constituido la base de la alimentación de todas las civilizaciones debido a sus características nutritivas. Es un alimento básico que forma parte de la dieta tradicional.

Está elaborada fundamentalmente con harina de cereales, sal y agua. La mezcla, en la mayoría de las ocasiones, suele contener levaduras para que fermente la masa y sea más esponjosa y tierna.

El cereal más utilizado para la elaboración del pan es la harina de trigo. También se utiliza el centeno, la cebada, el maíz y el arroz. Existen muchos tipos de pan que pueden contener otros ingredientes, como grasas de diferentes tipos (tocino de cerdo o de vaca, mantequilla, aceite de oliva), huevos, azúcar, especias, frutas, frutas secas (como por ejemplo pasas), verduras (como cebollas), o semillas diversas.

- **Complementos dietéticos.**

Son preparados que contienen uno o más nutrientes (vitaminas, minerales y carbohidratos) y se toman además de la comida diaria. Es muy poco probable que mejoren la salud, a excepción de que se sufra alguna deficiencia, en concreto, de un determinado nutriente.

Los médicos no son partidarios de que rutinariamente se consuman vitaminas y minerales adicionales, bajo determinadas circunstancias, tales como el embarazo, sí lo recomiendan. También las enfermedades crónicas, o algunas otras, causan deficiencias. Los complementos deben tomarse única y exclusivamente con prescripción facultativa, posterior a un análisis de la sangre y un examen médico.

- **Morbilidad.**

Es la cantidad de individuos considerados enfermos o que son víctimas de enfermedad en un espacio y tiempo determinado. La morbilidad es un dato estadístico importante para comprender la evolución o retroceso de alguna enfermedad, las razones de su surgimiento y las posibles soluciones.

En el sentido de la epidemiología se puede ampliar al estudio y cuantificación de la presencia y efectos de alguna enfermedad en una población.

- **Gestante.**

Es la etapa del embarazo de la mujer, se inicia con la fecundación y termina con el parto.

- **Preclamsia.**

Es una complicación en el embarazo, asociada a la hipertensión inducida durante el embarazo, asociada a los elevados niveles de proteína en la orina.

- **Feto.**

Es una de las etapas del desarrollo del proceso embrionario, ya que cuando ocurre la fecundación de los óvulos, este se convierte en un embrión, cuando este sigue creciendo, ya se convierte en un feto, es en esta etapa cuando se forman los órganos y los tejidos, esto ocurre en el tercer mes de la vida uterina:

- **Naive Bayes.**

Es un modelo más simple de clasificación de redes bayesianas. En este caso la estructura de la red es fija y solo necesitamos aprender los parámetros (probabilidades). El fundamento principal del clasificador Naive Bayes es la suposición de que los atributos son independientes conocido el valor de la variable clase. A pesar de que

asumir esta suposición en el clasificador Naive Bayes (NB) es sin duda bastante fuerte y poco realista en la mayoría de los casos, se trata de uno de los clasificadores más utilizados. Además, diversos estudios demuestran que sus resultados son competitivos con otras técnicas (redes neuronales y árboles de decisión entre otras) en muchos problemas y que incluso las superan en algunos otros.

- **Algoritmo k2.**

El algoritmo K2 es uno de los más rápidos para aprendizaje en RB y puede utilizarse para problemas supervisados y no supervisados, pero depende del orden que se establece entre las variables. No siempre es posible obtener el orden, por ejemplo las posiciones de las secuencias genómicas no son intercambiables y no es fácil establecer a priori un orden total de importancia.

- **Algoritmo B.**

Es un algoritmo basado en la optimización de una métrica de calidad de redes bayesianas. Al igual que el algoritmo K2, este se basa en un esquema voraz para la construcción de una solución aproximada a partir de la red vacía de enlaces. Este algoritmo no impone la restricción de proporcionarle como entrada un orden específico entre variables.

- **Algoritmo HC.**

Es un algoritmo de ascensión de colinas (hill climbing) por el máximo gradiente basado en la definición de una vecindad. El algoritmo parte de una solución inicial, como podría ser la red vacía de enlaces, u otra cualquiera.

A partir de esta solución se calcula el nuevo valor de la métrica utilizada de todas de las soluciones (grafos) vecinos a la solución actual y nos quedamos con el

vecino que mejor valor de la métrica resulte. Estos algoritmos al igual que los anteriores, se aprovechan de la descomponibilidad de las métricas para recalcar solo las modificaciones que se realizan en los grafos vecinos definidos.

- **Algoritmo Tan.**

Este algoritmo constituye una extensión del clasificador Naive Bayes. La idea es construir una red bayesiana (algo) más compleja que el Naive Bayes pero donde se da un tratamiento especial a la variable clase; por tanto, se encuadra en la filosofía de aprender redes bayesianas orientadas a clasificación. Los autores pretenden mantener la simplicidad computacional del clasificador Naive Bayes pero intentando mejorar la tasa de acierto durante la clasificación. Para ello, en lugar de suponer todas las variables independientes (dada la clase), se admiten ciertas dependencias entre los atributos. En concreto, se supone que los atributos constituyen una red bayesiana con forma de árbol.

- **Algoritmo Ban.**

Se encuadra en la filosofía de TAN, es decir, aprender redes bayesianas orientadas a clasificación. En BAN (Bayesian Network Augmented Naive Bayes) se procede aprendiendo una red bayesiana para los atributos (excluyendo la clase) y posteriormente se aumenta el modelo añadiendo la variable clase C y aristas desde C hacia todos los atributos. Para aprender la red puede usarse cualquier algoritmo de aprendizaje de redes bayesianas.

Capítulo III:

Métodos y Materiales

3.1 Metodología de Investigación.

3.1.1 Nivel de investigación.

La metodología de investigación usada para el presente trabajo comprende las siguientes fases:

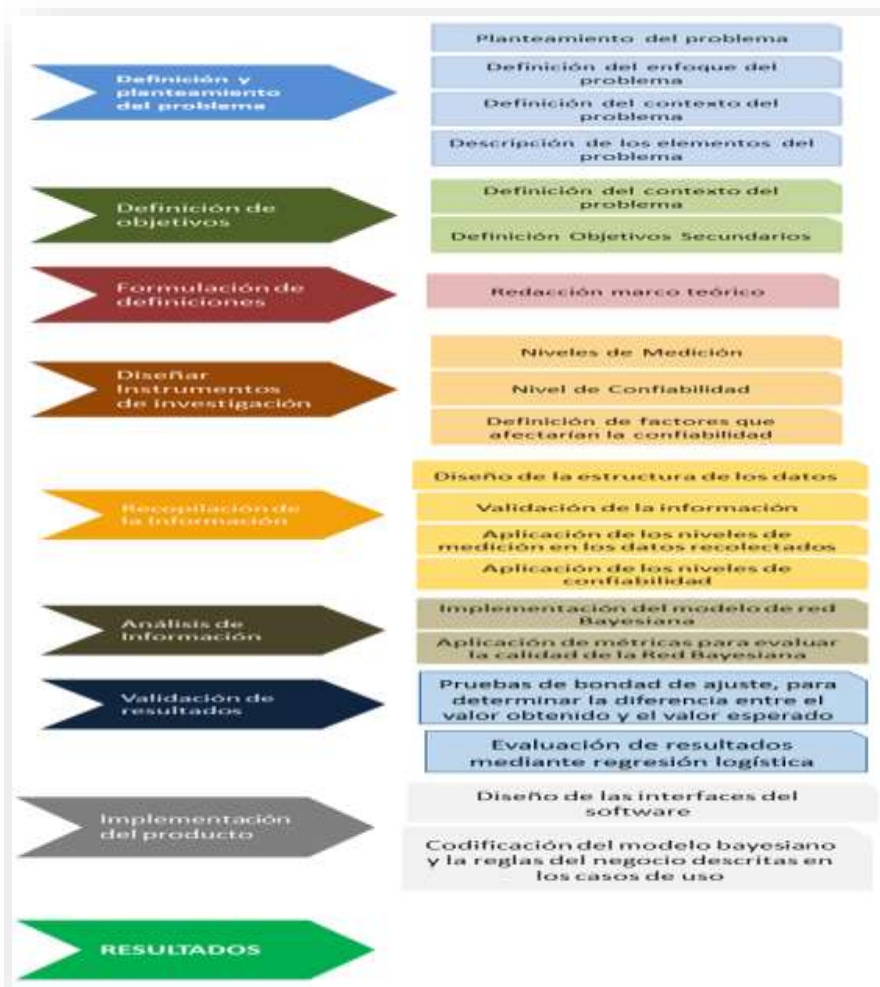


Figura 2. Metodología de Sampieri

3.1.2 Tipo de investigación.

El presente trabajo de investigación es de tipo descriptivo correlacional y predictivo, ya que buscamos la relación entre la variable del aspecto nutricional de la madre gestante en estudio, con su morbilidad neonatal.

3.1.3 Enfoque de la investigación.

- ***Cuantitativo.***

La investigación en curso tiene un enfoque cuantitativo, pues en la toma de los datos de la muestra se encuentran variables del aspecto nutricional de las madres gestantes que son cuantitativas continuas y cuantitativas discretas.

Esta información se recogió en base a la técnica de encuesta, y el cuestionario como instrumento. Además para la validación de la interfaz se utilizaron los CheckList de las madres gestantes.

3.1.4. Método de intervención tecnológica

En el Diagrama de bloques se representa el proceso que se siguió en el desarrollo de la investigación. En el siguiente cuadro indicamos el proceso que se planteó para el desarrollo de esta investigación.

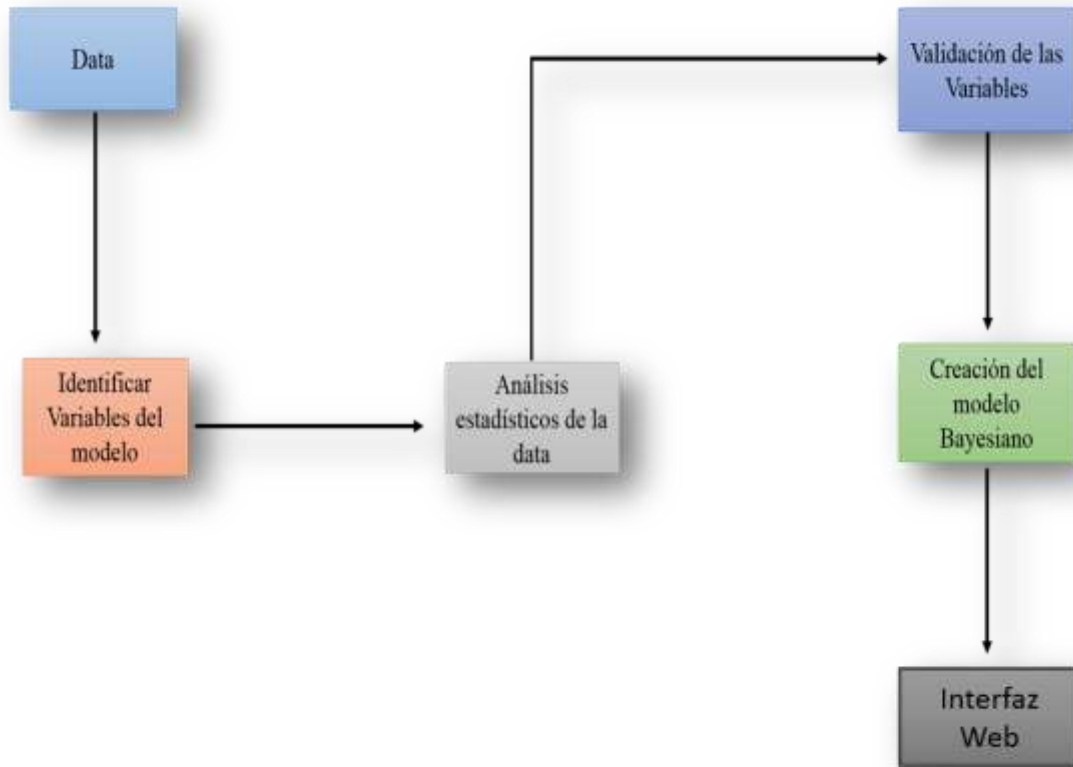


Figura 3. Diagrama de bloques

3.1.5 Población y muestra

A Población de estudio

En esta investigación se obtuvieron en 120 días, mediciones de madres gestantes registradas, en sus 13 variables dentro del proceso de control de la morbilidad del neonato, el cual constituye nuestra población de estudio.

B Determinación de la muestra

La muestra de la investigación ha sido seleccionado en forma probabilística del total de las 2 madres, usando la siguiente formula estadística:

$$n = \frac{N \times Z_{1-\frac{\alpha}{2}} \times S^2}{E^2(N-1) + Z^2 \times S^2}$$

Donde:

S=desviación estándar

N=población

E= error de precisión

Z=valor critico de tabla = 1.96

Tabla 4. Tabla de Población y muestra

Tamaño de la población	N	120
Error alfa	α	0.05
Nivel de confianza	1- α	0.95
Z de (1- α)	Z(1- α)	1.96
Desviación estándar	s	0.15
Precisión	E	0.05
Tamaño de la muestra	n=	27.01575263

La muestra queda constituida por 27 días de mediciones por las madres gestantes.

C Tipo de muestreo.

El tipo de muestreo es muestreo probabilístico, aleatorizado, el cual de los 120 días de mediciones por madres gestantes, todos los días tuvieron la probabilidad de salir seleccionadas en la muestra del estudio de investigación.

3.1.6 Presentación de la información.

El producto va a ser un modelo bayesiano que estará expresado en una herramienta, el entorno de desarrollo será un sistema web alimentada con datos

ya obtenidos (Datos procesados) y posteriormente alimentará la Base de datos del software Netica, que ayudará a las predicciones de Morbilidad del Neonato.

3.2 Materiales:

Para el siguiente proyecto, se contara con el siguiente Hardware:

Tabla 5. Descripción del Hardware

Hardware	
<i>Nombre</i>	<i>Descripción</i>
Procesador	AMD FX6300
Mainboard	MSI 970
Ram	8gb Kingston HyperX Fury
Disco Duro	SEAGATE Barracuda 1Tb
Monitor	Samsung 19.5" Led
Impresora	Epson Ecotank L455

- Software a utilizar:
 - Visual studio.net 2013
 - Sqlserver 2012
 - Netica 518
 - Excel 2013
 - Word 2013
 - Minitab 16
- Útiles de escritorio y otros
 - Cuadernillos para apuntes
 - 1 millar de Hojas Bond.
 - 3 lapiceros

3.3 Cronograma de actividades.

Tabla 6. Cronograma de Actividades

Nombre de tarea	Trabajo	Duración	Comienzo	Fin
Definición y planteamiento del problema	96 horas	5 días	lun 20/04/15	vie 24/04/15
<i>Nestor Ocaña Sudario</i>	48 horas		lun 20/04/15	vie 24/04/15
<i>Miguel Artica Chacón</i>	48 horas		lun 20/04/15	vie 24/04/15
Diagramas UML	24 horas	7.13 días	lun 27/04/15	Mié 29/04/15
<i>Miguel Artica Chacón</i>	24 horas		lun 27/04/15	mié 29/04/15
Definición de objetivo primario y objetivo secundario	16 horas	1 día	lun 04/05/15	lun 04/05/15
<i>Nestor Ocaña Sudario</i>	8 horas		lun 04/05/15	lun 04/05/15
<i>Miguel Artica Chacón</i>	8 horas		lun 04/05/15	lun 04/05/15
Analizar, preparar y validar los datos que serán usados por el modelo bayesiano	0 horas	6 días	jue 07/05/15	jue 14/05/15
<i>Miguel Artica Chacón</i>	0 horas		jue 07/05/15	jue 07/05/15
Definir los datos que serán usados por el modelo	0 horas	4 días	jue 07/05/15	mar 12/05/15
Depurar los datos y verificar la integridad de los mismos.	0 horas	3 días	mar 12/05/15	jue 14/05/15
Diseñar e implementar la Base de datos que será usada por el Modelo.	32 horas	4 días	mar 19/05/15	vie 22/05/15
<i>Miguel Artica Chacón</i>	32 horas		mar 19/05/15	mar 19/05/15
Creación de tablas SQL y Script para importación de datos.	32 horas	4 días	mar 19/05/15	vie 22/05/15
Redacción del marco teórico	50 horas	10 días	lun 11/05/15	vie 22/05/15
<i>Nestor Ocaña Sudario</i>	50 horas		lun 11/05/15	lun 11/05/15
Revisión de los resúmenes de los artículos relacionados al tema	8 horas	3 días	lun 11/05/15	mié 13/05/15

Revisión de conceptos relacionados a las redes Bayesianas	30 horas	2 días	jue 14/05/15	vie 15/05/15
Redacción del marco teórico	12 horas	5 días	lun 18/05/15	vie 22/05/15
Evaluación y selección del modelo Bayesiano	192 horas	18 días	lun 25/05/15	mié 17/06/15
<i>Miguel Artica Chacón</i>	<i>96 horas</i>		<i>lun 25/05/15</i>	<i>mar 09/06/15</i>
Revisión de los modelos bayesianos que mejor se adapte a nuestros procesos.	10 horas	5 días	lun 25/05/15	vie 29/05/15
Construcción de la estructura gráfica del modelo	40 horas	5 días	lun 01/06/15	vie 05/06/15
<i>Miguel Artica Chacón</i>	<i>40 horas</i>		<i>lun 01/06/15</i>	<i>vie 05/06/15</i>
Estimación de parámetros	24 horas	3 días	mar 09/06/15	jue 11/06/15
<i>Miguel Artica Chacón</i>	<i>24 horas</i>		<i>mar 09/06/15</i>	<i>jue 11/06/15</i>
Evaluación de bondad de ajuste	32 horas	4 días	vie 12/06/15	mié 17/06/15
<i>Nestor Ocaña Sudario</i>	<i>32 horas</i>		<i>vie 12/06/15</i>	<i>mié 17/06/15</i>
Diseño de la metodología de investigación	6 horas	1 día	vie 12/06/15	vie 12/06/15
<i>Nestor Ocaña Sudario</i>	<i>6 horas</i>		<i>vie 12/06/15</i>	<i>vie 12/06/15</i>
Selección de la población y muestra	6 horas	1 día	vie 12/06/15	vie 12/06/15
Implementación del software	256 horas	17 días	jue 18/06/15	vie 10/07/15
<i>Nestor Ocaña Sudario</i>	<i>64 horas</i>		<i>jue 18/06/15</i>	<i>lun 29/06/15</i>
<i>Miguel Artica Chacón</i>	<i>64 horas</i>		<i>jue 18/06/15</i>	<i>lun 29/06/15</i>
Codificación de las interfaces de usuario	16 horas	2 días	jue 18/06/15	vie 19/06/15
<i>Nestor Ocaña Sudario</i>	<i>16 horas</i>		<i>jue 18/06/15</i>	<i>vie 19/06/15</i>
Diseño e implementación de la arquitectura de Software	80 horas	10 días	lun 22/06/15	vie 03/07/15
<i>Miguel Artica Chacón</i>	<i>80 horas</i>		<i>lun 22/06/15</i>	<i>vie 03/07/15</i>

Integración del modelo Bayesiano con el Software	32 horas	4 días	mar 07/07/15	vie 10/07/15
<i>Miguel Artica Chacón</i>	<i>32 horas</i>		<i>mar 07/07/15</i>	<i>vie 10/07/15</i>
Pruebas de funcionalidad del Software	80 horas	5 días	lun 13/07/15	vie 17/07/15
<i>Nestor Ocaña Sudario</i>	<i>40 horas</i>		<i>lun 13/07/15</i>	<i>vie 17/07/15</i>
<i>Miguel Artica Chacón</i>	<i>40 horas</i>		<i>lun 13/07/15</i>	<i>vie 17/07/15</i>

3.4 Costos de la investigación.

- Presupuesto de los costos (recursos humanos, hardware y software)

Tabla 7. Presupuesto de costos y recursos humanos

Cargo	Costos S./
Líder de Proyecto	3500
Arquitecto	2500
Analista	3000
Tester	1000
Programador	2000
Especialista Redes bayesianas	800
Asesor estadístico	800
Asesor de proyecto	800
Total	11400

Tabla 8. Presupuesto de hardware

Descripción	Costos S./
Pc	2500
Impresora multifuncional	800
Cámara fotográfica	800
Total	4100

Tabla 9. Presupuesto de costo, software

Producto	Costo S./
Visual studio.net 2013	2167.45
Sqlserver 2012	2589.75
Excel 2013	398.65
Word 2013	398.65
Netica	1842.75
Total	7397.25

CAPÍTULO IV:

INGENIERÍA DE LA PROPUESTA

4.1 Proceso de pronóstico de la gestante

El proceso de la madre gestante con respecto a la morbilidad del neonato sigue varios pasos, entre las cuales figura el control mensual de su estado nutricional y su proceso de salud del neonato, durante los 9 meses de gestación. El aspecto nutricional y el estado de salud de la madre influye mucho, con respecto a la morbilidad del neonato, es por eso que se usan los indicadores de nutrición para tener un mejor control y pronóstico del estado de salud de la madre gestante y la morbilidad del neonato.

4.1.1 Indicadores de pronóstico

Los indicadores que se usan en la investigación, para llevar el control de la madre gestante en su primer trimestre son: trigo, harina de maíz, de avena, harina blanca, arroz integral y blanco, sémola, lechuga, espinaca, brócoli, zanahorias, calabaza, papas, maíz, garbanzos, lentejas, plátanos, manzana, naranja, arándanos, frutas secas, pera, kiwi, mangos, uva, melón, piña, aceite de oliva, aceite de maíz, aceite, de girasol, aceite de soja, aceite de almendras, manteca, mantequilla, margarina, leches enteras y descremadas, postres y lácteos, quesos, yogures, carne de vaca, carne de cerdo, carne de cordero, mariscos, pavo, pollo, pescados, huevos, otros, ácido fólico, otras vitaminas, agua, cafeínas, gaseosa, vino, alcohol, alimentos procesados, IMC, peso, desayuno, media mañana, almuerzo, media tarde, cena, noche, panes, azúcar, sal, condimentos, infección puerperal, morbilidad, muerte fetal, tamaño feto, preeclamsia.

Estas variables y el valor de las mismas varían en cuanto al régimen que deberá asignarse a cada madre gestante en su proceso de embarazo.

4.1.2 Proceso de pronóstico

El proceso de pronóstico se ejecuta de acuerdo a las fechas de registro de las madres gestante que se hace en el aplicativo web y las indicaciones que hace el especialista con respecto a su control mensual.

El llenado de la información, con los datos registrados durante el proceso de gestación de las madres, serán procesadas por el especialista.

La interfaz web ayudará al especialista en el pronóstico de su proceso de gestación, el cual proveerá la información necesaria para tomar una decisión de la madre gestante con respecto a su estado nutricional y morbilidad del neonato, durante su proceso de embarazo.

Si en la interfaz el especialista visualiza las carencias que tuviera la madre en su alimentación o algún desbalance en cuanto al cuidado que debe tener en su embarazo, el especialista deberá proporcionarle una nueva indicación y medicación nutricional para su mejor control de embarazo.

4.1.3 Dificultades

Las dificultad que se encontró con respecto al proceso del registro de los datos de la madre gestante es la falta de compromiso con respecto al llenado de los datos, ya que muchas de las madres gestante no cuenta con una conexión a Internet o en algunos casos no pudieron registrar por diferentes motivos (falta de luz, se olvidaron de registrar en el sistema, entre otros), es por eso que sistema en algunos casos no genera un mejor pronóstico.

4.2 Diagramas De Interfaz de la Ingeniería

4.2.1 Casos de uso del negocio.

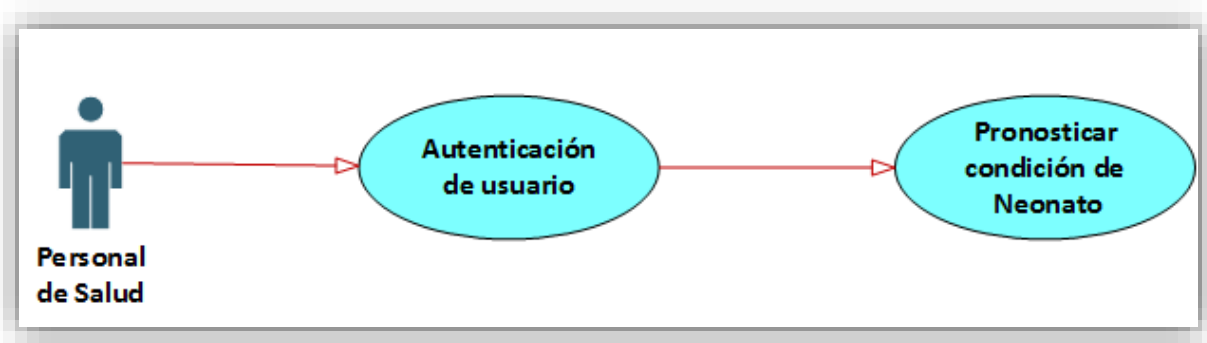


Figura4. Casos de uso del negocio - diseño de Modelo Bayesiano para el pronóstico de morbilidad del neonato

4.2.2 Casos de uso del sistema.

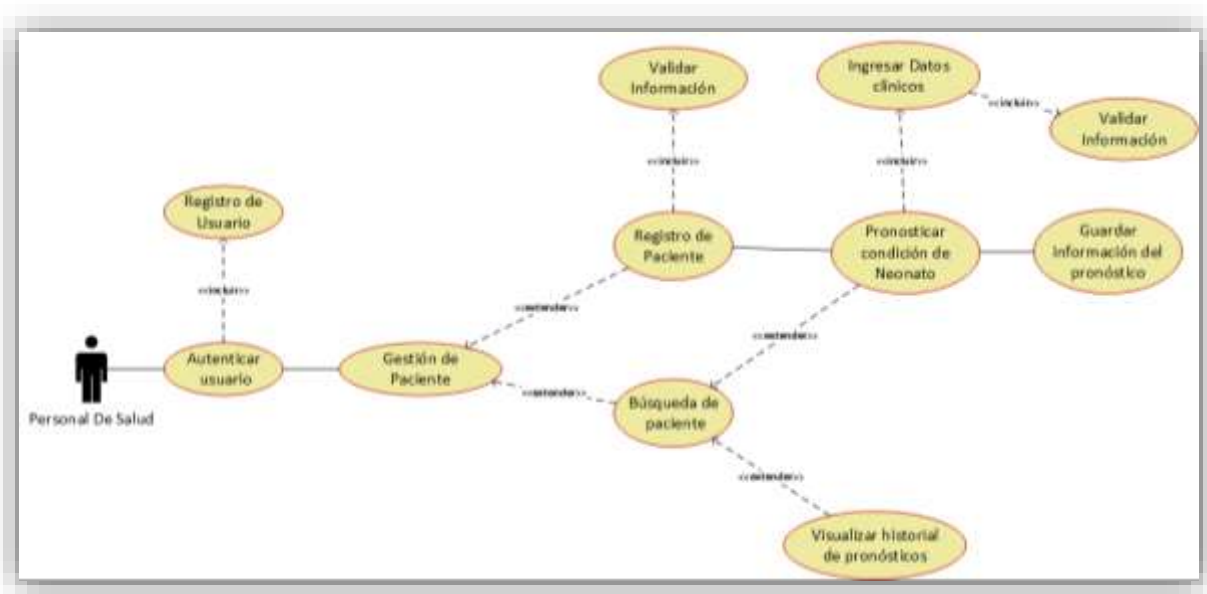


Figura 5. Casos de uso del sistema - diseño del Modelo Bayesiano para el pronóstico de morbilidad del neonato

4.2.3 Diagrama de Colaboración

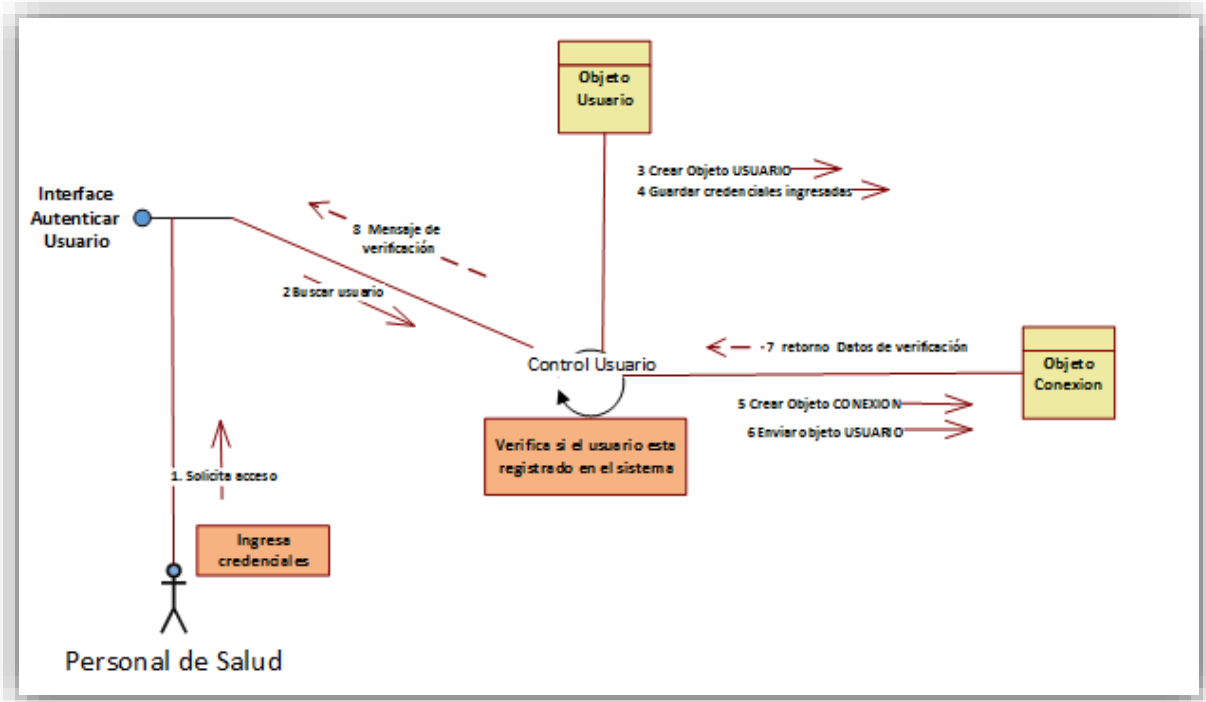


Figura6. Diagrama de Colaboración - Autenticación de usuario.

4.2.4 Diagrama de paquetes.

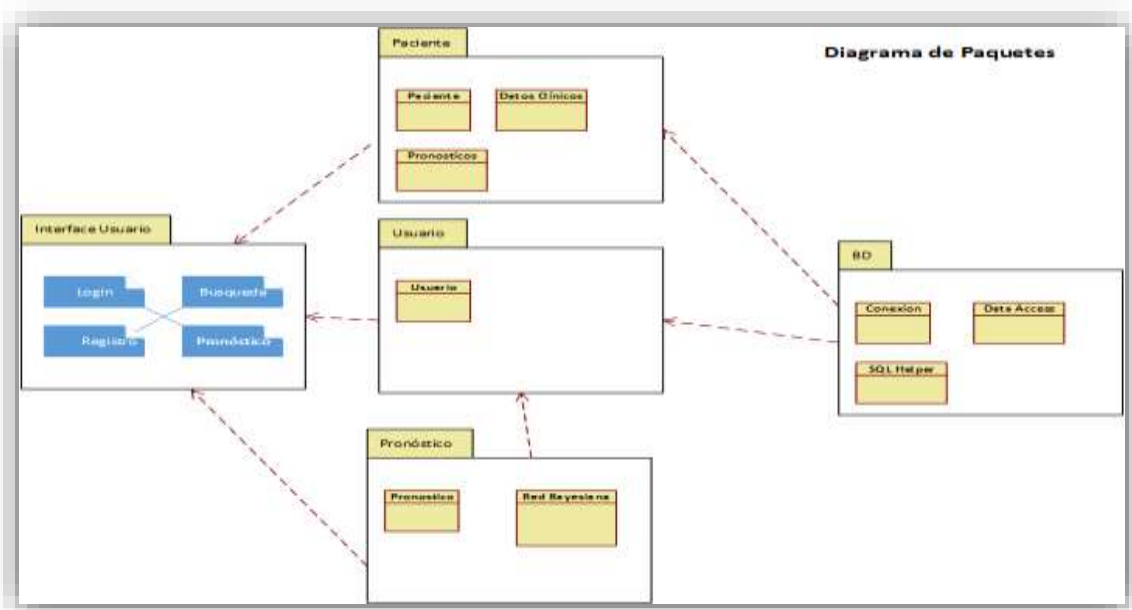


Figura7. Diagrama de paquetes

4.3 Restricciones del Modelo Bayesiano

En la interfaz web se manejaron algunas restricciones en cuanto al manejo de algunas variables con las que se trabajaron, como por ejemplo, la variable edad que debería estar incluido como un dato más dentro del sistema, sin embargo, se le registra como una paciente, dependiendo de la edad que tenga la paciente, porque para el uso de esta variable también.

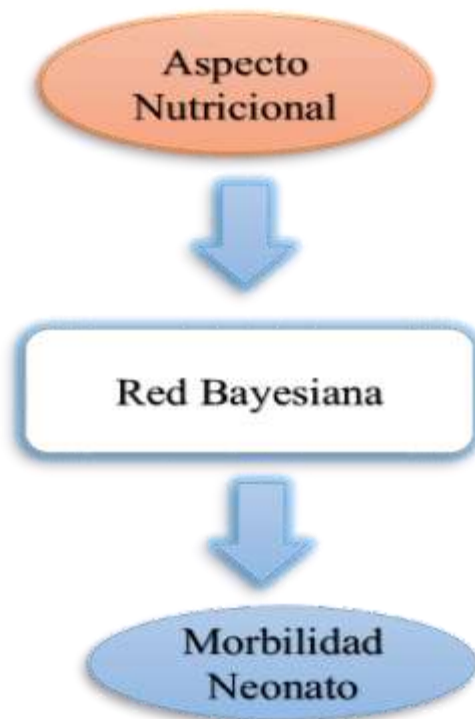


Figura 8. Diagrama de morbilidad neonatal

4.4 Variables del Modelo

4.4.1 Identificación de las variables del Modelo

Las variables del modelo que utiliza el algoritmo dentro de la interfaz, es como se muestra a continuación a acuerdo a las dimensiones de la interfaz:

Dimensión 1 = Cereales.

X1 Trigo

X2 Harina de maíz, de avena, harina blanca

X3 Arroz integral y blanco

X4 Sémola

Dimensión 2 = Verduras y hortalizas

X5 Lechuga

X6 Espinaca

X7 Brócoli

X8 Zanahorias

X9 Calabaza

X10 Papas

X11 Maíz

X12 Garbanzos

X13 Lentejas

Dimensión 3 = Frutas

X14 Plátanos

X15 Manzana

X16 Naranja

X17 Arándanos, frutas secas

X18 Pera

X19 Kiwi

X20 Mangos

X21 Uva

X22 Melón, piña

Dimensión 4 = Aceites y grasas

X23 Aceite de oliva

X24 Aceite de maíz

X25 Aceite de girasol

X26 Aceite de soja

X27 Aceite de almendras

X28.Manteca

X29 Mantequilla

X30 Margarina

Dimensión 5= Productos lácteos

X31 Leches enteras y descremadas

X32 Postres y lácteos

X33 Quesos

X34 Yogures

Dimensión 6 = Carnes, legumbres y huevos

X35 Carne de vaca

X36 Carne de cerdo

X37 Carne de cordero

X38 Mariscos

X39 Pavo

X40 Pollo

X41 Pescados

X42 Huevos

X43 Otros

Dimensión 7 = Complementos vitamínicos

X44 Ácido fólico

X45 Otras vitaminas

D 8 = Agua

X46 Agua

D 9 = Descalsificante

X 47 Cafeínas

X48 Gaseosa

X49 Vino

X50 Alcohol

X51 Alimentos procesados

D 10 = Tipo de alimentación

X52 Desayuno

X53 Media mañana

X54 Almuerzo

X55 Media tarde

X56 Cena

X57 Noche

D 11 = Panes

X58 Panes

D 12 = Azúcar

X59 Azúcar

D 13 = Sal

X60 Sal

D14 = Condimentos

X61Condimentos

Y = Morbilidad

Z = Modelo bayesiano

4.4.2 Variables de estado del Modelo

D1 = Cereales

D2 = Verduras y hortalizas

D3 = Frutas

D4 = Aceites y grasas

D5= Productos lácteos

D6 = Carnes, legumbres y huevos

D7 = Complementos vitamínicos

D9 = Descalsificante

D10 = Tipo de alimentación

D11 = Panes

D12 = Azúcar

D14 = Condimentos

4.5 Topología de la Red Bayesiana

4.5.1 Naive Bayes

Es un modelo más simple de clasificación de redes bayesianas. En este caso la estructura de la red es fija y solo necesitamos aprender los parámetros (probabilidades). El fundamento principal del clasificador Naive Bayes es la suposición de que los atributos son independientes conocido el valor de la variable clase. A pesar de que asumir esta suposición en el clasificador Naive Bayes (NB)

es sin duda bastante fuerte y poco realista en la mayoría de los casos, se trata de uno de los clasificadores más utilizados. Además, diversos estudios demuestran que sus resultados son competitivos con otras técnicas (redes neuronales y árboles de decisión entre otras) en muchos problemas y que incluso las superan en algunos otros.

Debido a la hipótesis de independencia usada en el Naive Bayes, la expresión para obtener la hipótesis MAP queda como sigue:

$$C_{MAP} = \arg_{c \in \Omega_c} \max p(A_1, \dots, A_n / c) p(c) = \arg_{c \in \Omega_c} \max p(c) \prod p(A_i / c)$$

Es decir, la tabla de probabilidad $p(A_1, \dots, A_n / c)$ ha sido factorizada como el producto de n tablas que solo involucran dos variables. Por tanto, los parámetros que tenemos que estimar son $p(A_i / c)$ para cada atributo y la probabilidad a priori de la variable clase $p(c)$

Veamos cómo hacerlo dependiendo de que el atributo $p(A_i)$ sea discreto o continuo.

Atributos discretos. En este caso la estimación de la probabilidad condicional se basa en las frecuencias de aparición que obtendremos en la base de datos. Así, si llamamos $n(x_i, Pa(x_i))$ al número de registros de la base de datos en que la variable X_i toma el valor x_i y los valores de X_i ($Pa(X_i)$) toman la configuración denotada por $Pa(X_i)$, entonces la forma más simple de estimar $P(x_i / Pa(x_i))$ es:

$$P(x_i / Pa(x_i)) = \frac{n(x_i, Pa(x_i))}{n(Pa(x_i))}$$

Es decir, el número de casos favorable dividido por el número de casos totales.

Esta técnica se conoce como estimación por máxima verosimilitud y tiene como desventajas que necesita una muestra de gran tamaño y que sobre ajusta a los datos. Existen otros estimadores más complejos que palian estos problemas, entre ellos citaremos el estimador basado en la ley de la sucesión de Laplace:

$$P(x_i/Pa(x_i)) = \frac{n(x_i, Pa(x_i)) + 1}{n(Pa(x_i)) + |\Omega_{x_i}|}$$

El número de casos favorable más uno dividido por el número de casos totales más el número de valores posibles. Junto a otros estimadores, como el m estimado. Con pocos ejemplos, la probabilidad se corrige por la probabilidad uniforme a priori, es decir, uno dividido por el número de valores posibles. Con esta estimación lo que se pretende es que todas las configuraciones posibles tengan una mínima probabilidad, ya que con el estimador de máxima verosimilitud cualquier configuración que no esté presente en la base de datos tendrá la posibilidad cero.

Atributos continuos: en este caso, el clasificador de Naive Bayes supone que el atributo en cuestión sigue una distribución normal-, por tanto, lo único que tenemos que calcular (a partir de la base de datos) es la media μ y la desviación típica σ condicionales a cada valor de la variable clase.

$$P(A_i/c) \propto N(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(X - \mu)^2}{2\sigma^2}\right)$$

4.5.2 Nodos.

En una red bayesiana, los nodos son las que representan a las variables aleatorias (cereales, panes, sal, etc.), a su vez representan una entidad del mundo real y los arcos son que representan a las relaciones de influencia casual entre ellas.

4.6 Modelo probabilístico

4.6.1 Distribuciones de probabilidad

Se utilizó la distribución de Poisson para las variables cualitativas ordinarias y la distribución de T Student para la correlación de las variables.

Distribución de Poisson.

Es una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media, la probabilidad de que ocurra un determinado número de eventos durante cierto período de tiempo.

La función de masa o probabilidad de la distribución de Poisson es:

$$f(k, \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Donde:

- Es el número de ocurrencias del evento o fenómeno (la función nos da la probabilidad de que el evento suceda precisamente veces).
- Es un parámetro positivo que representa el número de veces que se espera que ocurra el fenómeno durante un intervalo dado. Por ejemplo, si el suceso estudiado tiene lugar en promedio 4 veces por minuto y estamos interesados en la probabilidad de que ocurra veces dentro de un intervalo de 10 minutos, usaremos un modelo de distribución de Poisson con $\lambda = 10 \times 4 = 40$.
- Es la base de los logaritmos naturales ($e = 2,71828..$)

T Student.

Es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño.

A la teoría de pequeñas muestras también se le llama teoría exacta del muestreo, ya que también la podemos utilizar con muestras aleatorias de tamaño grande.

Veremos un nuevo concepto necesario para poder entender la distribución T Student. Este concepto es "grados de libertad".

Para definir grados de libertad se hará referencia a la varianza maestral:

$$t = \rho_{D,y} \sqrt{\frac{n-2}{1-\rho_{D,y}^2}}$$

Propiedades de las distribuciones t:

- Cada curva t tiene forma de campana con centro en 0.
- Cada curva t, está más dispersa que la curva normal estándar.
- A medida que k aumenta, la dispersión de la curva t correspondiente disminuye.
- A medida que $k \rightarrow \infty$, la secuencia de curvas t se aproxima a la curva normal estándar.

La distribución de probabilidad de t se publicó por primera vez en 1908 en un artículo de W. S. Gosset. En esa época, Gosset era empleado de una

cervecería irlandesa que desaprobaba la publicación de investigaciones de sus empleados. Para evadir esta prohibición, publicó su trabajo en secreto bajo el nombre de "Student". En consecuencia, la distribución t normalmente se llama distribución t de Student, o simplemente distribución t.

4.6.2 Algoritmo utilizado

Naive Bayes

Es un modelo más simple de clasificación de redes bayesianas. En este caso la estructura de la red es fija y solo necesitamos aprender los parámetros (probabilidades). El fundamento principal del clasificador Naive Bayes es la suposición de que los atributos son independientes conocido el valor de la variable clase. A pesar de que asumir esta suposición en el clasificador Naive Bayes (NB) es sin duda bastante fuerte y poco realista en la mayoría de los casos, se trata de uno de los clasificadores más utilizados. Además, diversos estudios demuestran que sus resultados son competitivos con otras técnicas (redes neuronales y árboles de decisión entre otras) en muchos problemas y que incluso las superan en algunos otros. Como un ejemplo de problema en el que el clasificador NB se está mostrando como una de las técnicas más eficaces, podemos citar la lucha contra el correo basura o spam. Muchos lectores de correo incorporan este clasificador para etiquetar el correo no solicitado.

La hipótesis de independencia asumida por el clasificador NB da lugar a un modelo gráfico probabilístico en el que existe un único nodo raíz (la clase), y en la que todos los atributos son nodos hoja que tienen como único padre a la variable clase. Gráficamente la estructura sería como sigue:

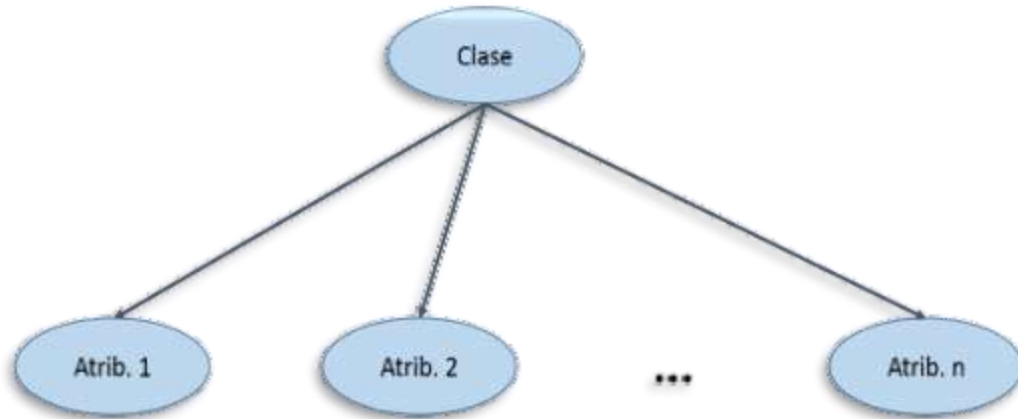


Figura 9. Topología de un clasificador Naive Bayes

Estimación de parámetros

Debido a la hipótesis de independencia usada en el Naive Bayes, la expresión para obtener la hipótesis MAP queda como sigue:

$$C_{MAP} = \arg_{c \in \Omega_c} \max p(A_1, \dots, A_n / c) p(c) = \arg_{c \in \Omega_c} \max p(c) \prod p(A_i / c)$$

Es decir, la tabla de probabilidad $p(A_1, \dots, A_n / c)$ ha sido factorizada como el producto de n tablas que solo involucran dos variables. Por tanto, los parámetros que tenemos que estimar son $p(A_i / c)$ para cada atributo y la probabilidad a priori de la variable clase $p(c)$.

Veamos cómo hacerlo dependiendo de que el atributo $p(A_i)$ sea discreto o continuo.

Atributos discretos. En este caso la estimación de la probabilidad condicional se basa en las frecuencias de aparición que obtendremos en la base de datos. Así, si llamamos $n(x_b, pa(x_i))$ al número de registros de la base de datos en que la variable X_i toma el valor x_i y los valores de X_i ($pa(X_i)$) toman la configuración denotada por $Pa(X_i)$, entonces la forma más simple de estimar $P(x_i / Pa(x_i))$ es:

$$P(x_i/Pa(x_i)) = \frac{n(x_i, Pa(x_i))}{n(Pa(x_i))}$$

Es decir, el número de caos favorable dividido por el número de casos totales.

Esta técnica se conoce como estimación por máxima verosimilitud y tiene como desventajas que necesita una muestra de gran tamaño y que sobre ajusta a los datos.

Existen otros estimadores más complejos que palian estos problemas, entre ellos citaremos el estimador basado en la ley de la sucesión de Laplace:

$$P(x_i/Pa(x_i)) = \frac{n(x_i, Pa(x_i)) + 1}{n(Pa(x_i)) + |\Omega_{x_i}|}$$

El número de casos favorable más uno dividido por el número de casos totales más el número de valores posibles. Junto a otros estimadores, como el m estimado. Con pocos ejemplos, la probabilidad se corrige por la probabilidad uniforme a priori, es decir, uno dividido por el número de valores posibles. Con esta estimación lo que se pretende es que todas las configuraciones posibles tengan una mínima probabilidad, ya que con el estimador de máxima verosimilitud cualquier configuración que no esté presente en la base de datos tendrá la posibilidad cero.

Atributos continuos: en este caso, el clasificador de Naive Bayes supone que el atributo en cuestión sigue una distribución normal-, por tanto, lo único que tenemos que calcular (a partir de la base de datos) es la media μ y la desviación típica σ condicionales a cada valor de la variable clase.

$$P(A_i/c) \propto N(\mu, \sigma) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left(-\frac{(X - \mu)^2}{2\sigma^2}\right)$$

Algoritmo K2

Es el primer algoritmo basado en búsqueda y optimización de una métrica bayesiana y ha sido fuente de inspiración para posteriores trabajos sobre el tema. Este algoritmo utiliza un esquema voraz en su búsqueda de soluciones candidatas cada vez mejores y parte de que las variables de entrada están ordenadas, de forma que los posibles padres de una variable aparecen en el orden antes que ella misma. Esta restricción es bastante fuerte pero fue estándar en el origen de los primeros trabajos sobre aprendizaje de redes bayesianas. El proporcionarle al algoritmo un orden entre las variables hace que éste tan “solo” tenga que buscar el mejor conjunto de padre posible de entre las variables predecesoras en el orden. La búsqueda de este conjunto se hace de forma voraz.

El algoritmo parte de que el conjunto de padres para cada es el conjunto vacío. Posteriormente, y siguiendo el orden establecido, pasa a procesar cada variable X_i , calcula la ganancia que se produce en la medida utilizada al introducir una variable X_j como padre de X_i entre todas sus predecesoras, esto es, para todo $j < i$ y se queda con la que produce la mejor ganancia. Desde un enfoque bayesiano, esta ganancia con variable X_i sin padres y una red bayesiana donde X_i tenga como padre X_j . Una vez fijado el primer padre de X_i , entonces se prosigue de igual forma, pero esta vez midiendo la razón de una red bayesiana con X_j y el primer padre fijado y una red bayesiana donde se le introduce un segundo padre no insertado previamente.

ALGORITMO K2 (X: nodos (variables ordenadas), D: Datos)

Fase de Iniciación:

PARA CADA X_i (i=0 hasta n)

Pa (X_i) = conjunto vacío

FIN PARA CADA

Fase Iterativa:

PARA CADA X_i (i=0 hasta n)

Ok:= true

MIENTRAS ok

Sea X_j el nodo tal que $j < i$ y X_j no pertenezca a Pa (X_i) que maximiza

$$f_i(X_i | Pa(X_i) \cup X_j : D)$$

SI $f_i(X_i | Pa(X_i) \cup X_j : D) > f_i(X_i | Pa(X_i) : D)$ **ENTONCES** Pa(X_i) := Pa(X_i) \cup X_j

EN CASO CONTRARIO OK: = false

FIN MIENTRAS

FIN PARA CADA

FIN ALGORITMO

Figura 10. Algoritmo K2

Algoritmo B

Este algoritmo es también de los primeros en aparecer en la literatura especializada en el tema, y también está basado en la optimización de una métrica de calidad de redes bayesianas. Al igual que el algoritmo K2, este se basa en un esquema voraz para la construcción de una solución aproximada a partir de la red vacía de enlaces (cada nodo posee inicialmente un conjunto vacío de padres). A diferencia del algoritmo previo este algoritmo no impone la restricción de proporcionarle como entrada un orden específico entre las variables.

ALGORITMO B (X: nodos (variables), D: Datos)

Fase de Iniciación:

PARA CADA X_i (i=0 hasta n)

Pa (X_i) = conjunto vacío

FIN PARA CADA

Fase Iterativa:

Ok:= true

MIENTRAS ok

Sea $X_j \rightarrow X_i$ el enlace que maximiza (de todos los enlaces que no formen un ciclo dirigido acíclico y previamente no incluido) la medida:

$$f_i(X_i | Pa(X_i) \cup X_j; D) - f_i(X_i | Pa(X_i); D)$$

SI $f_i(X_i | Pa(X_i) \cup X_j; D) - f_i(X_i | Pa(X_i); D)$ **ENTONCES** $Pa(X_i) := Pa(X_i) \cup X_j$

EN CASO CONTRARIO OK: = false

FIN MIENTRAS

FIN ALGORITMO

Figura11. Pseudocódigo del Algoritmo B

Este algoritmo trata de introducir el arco de mayor ganancia representada con respecto a la red anterior y a la métrica utilizada. El algoritmo se detiene cuando la inclusión de un arco no representa ninguna ganancia.

Algoritmo HC

Es un algoritmo de ascensión de colinas (hill climbing) por el máximo gradiente basado en la definición de una vecindad. El algoritmo parte de una solución inicial, como podría ser la red vacía de enlaces, u otra cualquiera (por ejemplo, en tareas de clasificación se podría inicializar con una estructura Naive Bayes).

A partir de esta solución se calcula el nuevo valor de la métrica utilizada de todas las soluciones (grafos) vecinos a la solución actual y nos quedamos con el vecino que

mejor valor de la métrica resulte. Estos algoritmos al igual que los anteriores, se aprovechan de la descomponibilidad de las métricas para recalcular solo las modificaciones que se realizan en los grafos vecinos definidos.

La vecindad clásica que se maneja en este algoritmo es la siguiente: dado un grafo G (solución actual) se denomina grafos vecindad G' a aquellos resultantes de incluir un solo arco a G o borrar un solo arco presente en G o invertir la dirección de un arco presente en G , todo ello sin incluir ciclos dirigidos en G' . El algoritmo parará cuando no exista ningún vecino que puede mejorar la solución actual (óptimo local).

Algoritmo TAN

TAN (Tree Augmented Naive Bayes) constituye una extensión del clasificador Naive Bayes. La idea es construir una red bayesiana (algo) más compleja que el Naive Bayes pero donde se da un tratamiento especial a la variable clase; por tanto, se encuadra en la filosofía de aprender redes bayesianas orientadas a clasificación. Con TAN, los autores pretenden mantener la simplicidad computacional del clasificador Naive Bayes pero intentando mejorar la tasa de acierto durante la clasificación. Para ello, en lugar de suponer todas las variables independientes (dada la clase), se admiten ciertas dependencias entre los atributos. En concreto, se supone que los atributos constituyen una red bayesiana con forma de árbol.

La ventaja de restringir la topología de la red (entre los atributos) a un árbol, es que esta estructura puede aprenderse eficientemente. Así, los autores de TAN proponen usar una ligera modificación del algoritmo de Chow y Liu para realizar este aprendizaje. Este algoritmo está basado en el concepto de información mutua:

$$I_p(X, Y/C) = \sum_{x, y, c} P(x, y, c) \log \frac{P(x, y, c)}{P(x/c)P(y/c)}$$

Que mide la cantidad de información que la variable Y nos proporciona sobre la variable X supuesto que el valor de la variable clase (C) es conocido.

Tras aprender la estructura de árbol entre los atributos, el algoritmo TAN añade la variable clase y la hace padre de todos ellos. Este algoritmo tiene dos claras ventajas: un bajo coste computacional $O(n^2 \cdot N)$, siendo N el número de instancias en el conjunto de entrenamiento y n el número de variables; y, en segundo lugar, asegura que la estructura de red obtenida es la máxima verosimilitud del conjunto de todas las posibles estructuras TAN.

ALGORITMO TAN

1. Calcular $I_p(A_i, A_j / C)$ para cada par de atributos ($i \neq j$).
2. Crear un grafo no dirigido con todos los atributos como conjunto de nodos y añadir aristas entre cada par de nodos.
3. Asociar a cada arista (i, j) del grafo el peso $I_p(A_i, A_j / C)$.
4. Construir un árbol expandido de máximo peso a partir del grafo anterior.
5. Elegir un nodo cualquiera del árbol anterior como raíz y direccionar a partir de él, el resto de aristas.
6. Añadir la variable clase C y el conjunto de aristas dirigidas $(C \rightarrow A_i)$ para todo atributo A_i .
7. Devolver el modelo TAN obtenido.

FIN ALGORITMO

Figura 12. Pseudocódigo Algoritmo TAN

Finalmente, indicar que existen distintas variaciones de este algoritmo; entre otras, podemos citar las siguientes:

- Métodos de aprendizaje de redes bayesianas que parten de la estructura del Naive Bayes como estado inicial y, progresivamente añaden arcos hasta llegar a una estructura de TAN. Denotaremos esta variante como TAN_{i-m} , siendo n el método de aprendizaje.

- Métodos que simultáneamente al proceso de construcción realizan una selección de variables (*sTAN*).
- Métodos que distribuyen los atributos en un conjunto de árboles, un bosque, en lugar de un único árbol (FAN, de Forest Augmented Naive Bayes).

Algoritmo BAN

El algoritmo se encuadra en la filosofía de TAN, es decir, aprender redes bayesianas orientadas a clasificación. En BAN (Bayesian Network Augmented Naive Bayes) se procede aprendiendo una red bayesiana para los atributos (excluyendo la clase) y posteriormente se aumenta el modelo añadiendo la variable clase C y aristas desde C hacia todos los atributos. Para aprender la red puede usarse cualquier algoritmo de aprendizaje de redes bayesianas.

Al igual que ocurre con TAN, existe la posibilidad de iniciar la estructura de red como un Naive Bayes y a partir de ahí lanzar un algoritmo de aprendizaje que vaya añadiendo arcos. A estos modelos los notaremos como BAN_{i-m} , siendo m el algoritmo de aprendizaje de redes bayesianas utilizado.

Algoritmo EM

Supongamos que en este caso que la estructura G de una red bayesiana esta perfectamente definida y que por tanto tan solo tendremos que estimar las tablas de distribución condicional asociadas a cada nodo o variable, como ocurre por ejemplo en el clasificador Naive Bayes. En este caso se puede utilizar el algoritmo EM (expectation Maximization). El algoritmo consta de dos etapas iterativas: en la primera de ellas se calculan las esperanzas de los estadísticos necesarios (frecuencias

esperadas N_{ijk}) a partir del modelo de la etapa actual; y el segundo paso es la fase de maximización consistente en reestimar los parámetros a partir de las frecuencias esperadas calculadas en el paso previo. El proceso se reitera hasta su convergencia, la cual esta teóricamente garantizada.

ALGORITMO EM ($B = (G, \Theta)$): Red Bayesiana, D: Datos)

Fase de Inicialización:

Inicializar el conjunto de parámetros $\Theta^{(0)}$

Inicializar contador de etapas $\Theta = 0$

Fase Iterativa:

MIENTRAS no convergencia

Paso E: etapa de cálculo de Esperanzas $E[N_{ijk} / \Theta^e]$

Paso M: etapa de maximización

$$\Theta_{ijk}^{(e+1)} = E[N_{ijk} / \Theta^e] / E[N_{ij} / \Theta^e]$$

FIN MIENTRAS

FIN ALGORITMO

Figura 13. Pseudocódigo Algoritmo EM

4.6.3 Análisis Estadístico

Validación hipótesis específica aceites y grasas.

Con el Pronóstico del Aspecto Nutricional apoyado por un Modelo Bayesiano y su relación en el pronóstico de Morbilidad del Neonato de la gestante.

Tabla 10. Correlación de la variable aceites y grasas

			Aceites y grasas	Morbilidad
Rho de Spearman	Aceites y grasas	Coeficiente de correlación	1,000	,679**
		Sig. (bilateral)	.	,008
		N	14	14
	Morbilidad	Coeficiente de correlación	,679**	1,000
		Sig. (bilateral)	,008	.
		N	14	14

Hipótesis Nula

H_0 = Si se aplica el aspecto nutricional aceites y grasas apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

H_a = Si se aplica el aspecto nutricional aceites y grasas apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=24$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D,y} = 0 \\ H_1 : \rho_{D,y} \neq 0 \end{cases}$$

$$t = \rho_{D_1y} \sqrt{\frac{n-2}{1-\rho_{D_1y}^2}}$$

$$t = 0.679 \sqrt{\frac{14-2}{1-(0.679)^2}}$$

$$t = 3.204$$

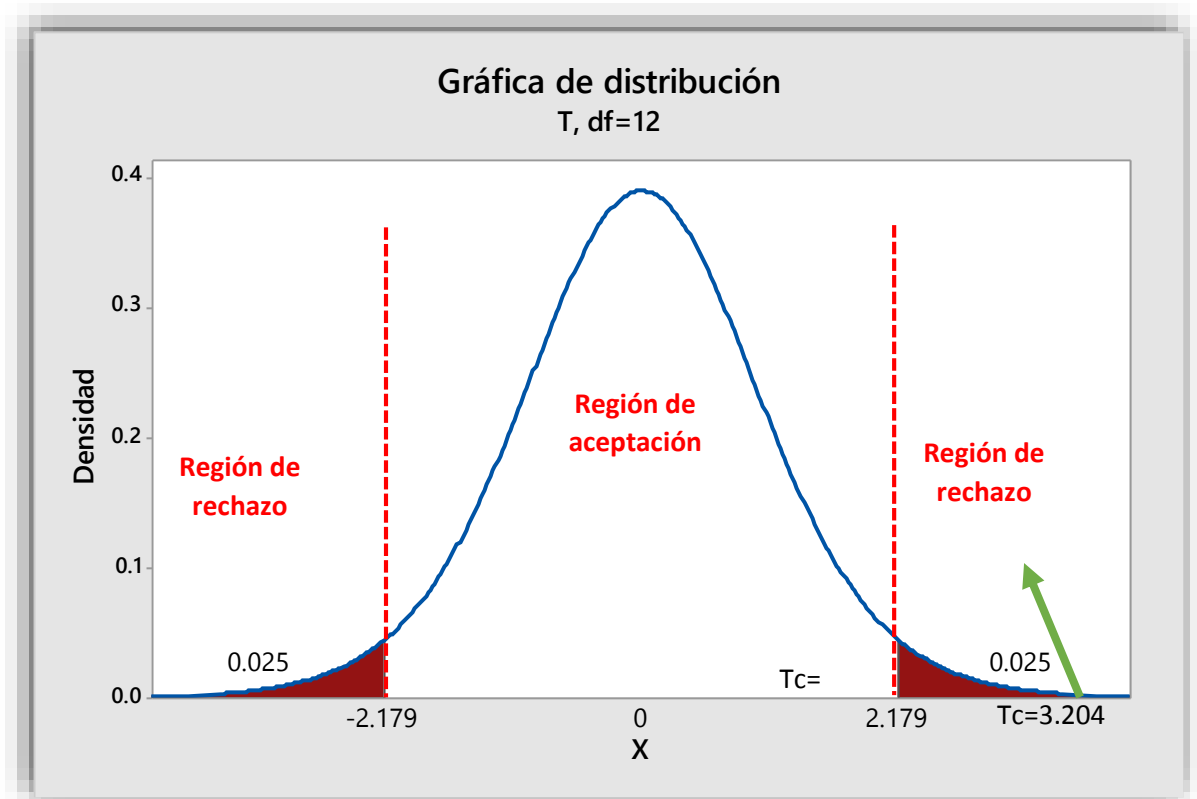


Figura14. Gráfica de la hipótesis aceites y grasas

Interpretación

Como $t=3.204 > t_c=2.179$ entonces rechazamos la hipótesis nula H_0 y nos quedamos con la hipótesis alterna H_a , en el nivel de significancia de $\alpha = 5\%$. El cual afirma la hipótesis alterna H_a . que dice: Si se aplica el aspecto nutricional aceites y grasas apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Validación Prueba de hipótesis agua

Tabla 11. Correlación de la variable agua

			Agua	Morbilidad
Rho de Spearman	Agua	Coefficiente de correlación	1,000	-,394
		Sig. (bilateral)	.	,163
		N	14	14
	Morbilidad	Coefficiente de correlación	-,394	1,000
		Sig. (bilateral)	,163	.
		N	14	14

Hipótesis Nula

H_0 = Si se aplica el aspecto nutricional agua apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

H_a = Si se aplica el aspecto nutricional agua apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D_1,y} = 0 \\ H_1 : \rho_{D_1,y} \neq 0 \end{cases}$$

$$t = \rho_{D_1,y} \sqrt{\frac{n-2}{1-\rho_{D_1,y}^2}}$$

$$t = -0.394 \sqrt{\frac{14-2}{1-(-0.394)^2}}$$

$$t = -1.485$$

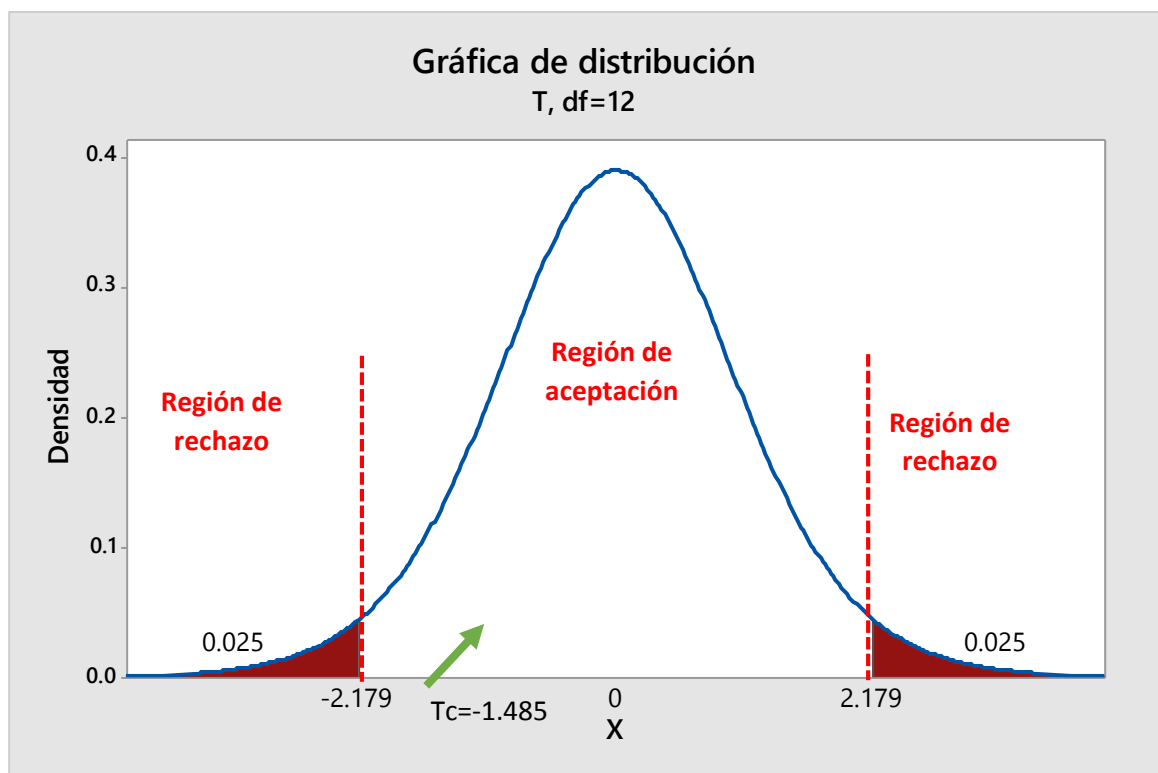


Figura 15. Gráfica de la hipótesis agua

Interpretación

Como $t = -1.485 < t_c = 2.179$ entonces aceptamos la hipótesis nula H_0 y nos quedamos con la hipótesis alterna H_a , en el nivel de significancia de $\alpha = 5\%$. El cual afirma la hipótesis alterna H_a , que dice: Si se aplica el aspecto nutricional Agua apoyado por

un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Prueba de hipótesis azúcar

Tabla 12. Correlación de la variable azúcar

			azúcar	Morbilidad
Rho de Spearman	Azúcar	Coeficiente de correlación	1,000	,070
		Sig. (bilateral)	.	,813
		N	14	14
	Morbilidad	Coeficiente de correlación	,070	1,000
		Sig. (bilateral)	,813	.
		N	14	14

Hipótesis Nula

H_0 = Si se aplica el aspecto nutricional azúcar apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

H_a = Si se aplica el aspecto nutricional azúcar apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D_1,y} = 0 \\ H_1 : \rho_{D_1,y} \neq 0 \end{cases}$$

$$t = \rho_{D_1,y} \sqrt{\frac{n-2}{1-\rho_{D_1,y}^2}}$$

$$t = 0.070 \sqrt{\frac{14-2}{1-(-0.070)^2}}$$

$$t = 0.243$$

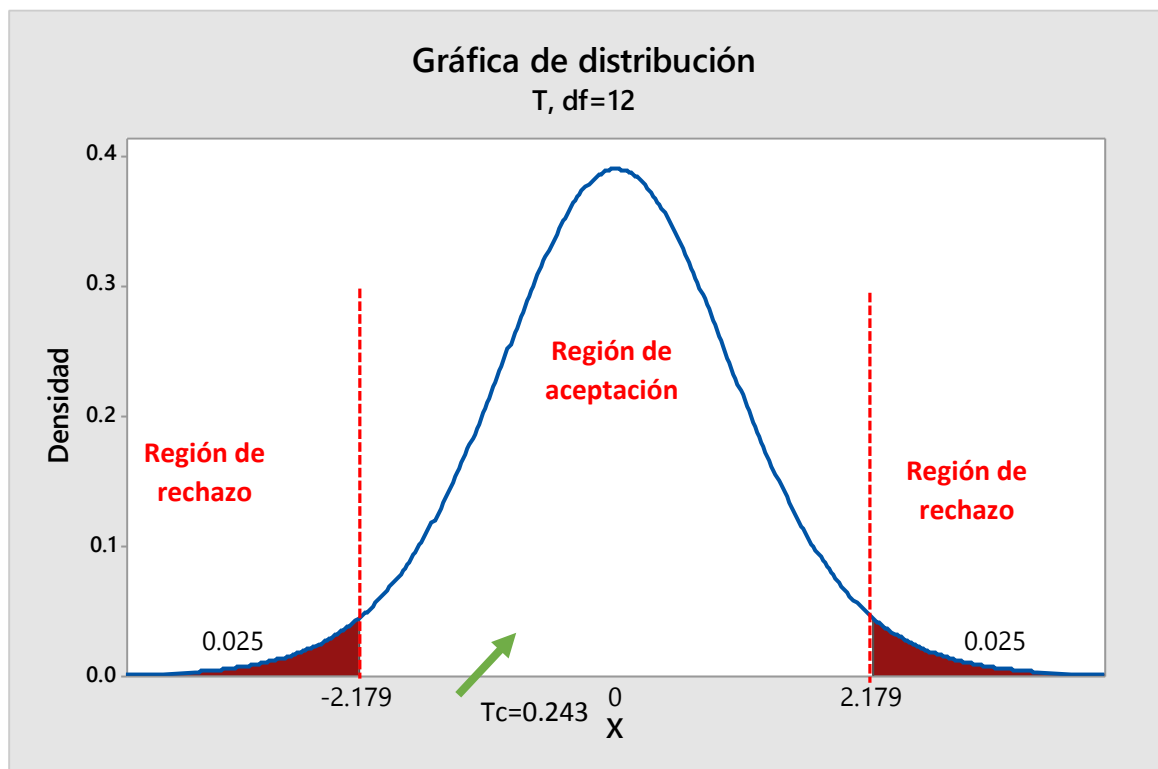


Figura 16. Grafica de la hipótesis Azúcar

Interpretación

En el grafico se observa que el valor de t calculado es $t=0.243$ que por ser menor al t_c (t critico) $t_c=2.179$ entonces aceptamos la hipótesis nula H_0 en el nivel de significancia $\alpha =5\%$ el cual afirma la hipótesis nula H_0 . Que dice: Si se aplica el

aspecto nutricional azúcar apoyado por un Modelo Bayesiano entonces no existe correlación significativa con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante. Cabe rescatar que el valor Rho de Spearman, es 0.070, que significa que hay una correlación muy débil entre estas 2 variables y es por esto que se acepta la hipótesis nula.

Prueba de hipótesis carnes y legumbres

Tabla 13. Correlación de la variable carnes y legumbres

		Carnes, legumbres y huevos	Morbilidad
Rho de Spearman	Carnes, Legumbres y	Coeficiente de correlación	1,000
	Huevos	Sig. (bilateral)	,175
		N	,549
			14
	Morbilidad	Coeficiente de correlación	,175
		Sig. (bilateral)	,549
		N	14

Hipótesis Nula

H_0 = Si se aplica el aspecto nutricional carnes y legumbres apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

H_a = Si se aplica el aspecto nutricional carnes y legumbres apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D,y} = 0 \\ H_1 : \rho_{D,y} \neq 0 \end{cases}$$

$$t = \rho_{D,y} \sqrt{\frac{n-2}{1-\rho_{D,y}^2}}$$

$$t = 0.175 \sqrt{\frac{14-2}{1-(0.175)^2}}$$

$$t = 0.616$$

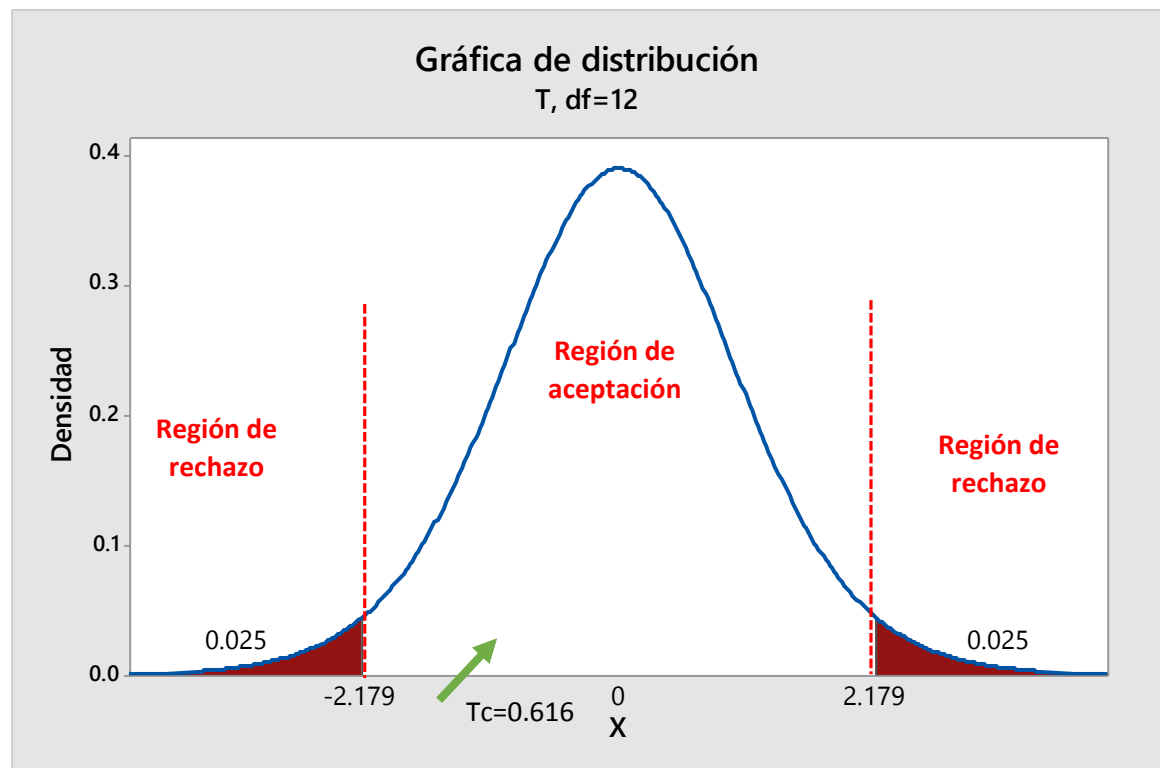


Figura17.Grafica de la hipótesis Carnes y legumbres

Interpretación

Como $t=0.616 < t_c=2.179$ entonces aceptamos la hipótesis nula H_0 y nos quedamos con la hipótesis alterna H_a , en el nivel de significancia de $\alpha =5\%$. El cual afirma la hipótesis alterna H_a , que dice: Si se aplica el aspecto nutricional carnes y legumbres apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Prueba de hipótesis cereales

Tabla 14. Correlación de la variable cereal

			cereales	Morbilidad
Rho de Spearman	Cereales	Coeficiente de correlación	1,000	,305
		Sig. (bilateral)	.	,289
		N	14	14
	Morbilidad	Coeficiente de correlación	,305	1,000
		Sig. (bilateral)	,289	.
		N	14	14

Hipótesis Nula

H_0 = Si se aplica el aspecto nutricional cereales apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

H_a = Si se aplica el aspecto nutricional cereales apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D,y} = 0 \\ H_1 : \rho_{D,y} \neq 0 \end{cases}$$

$$t = \rho_{D,y} \sqrt{\frac{n-2}{1-\rho_{D,y}^2}}$$

$$t = 0.305 \sqrt{\frac{14-2}{1-(0.305)^2}}$$

$$t = 1.109$$

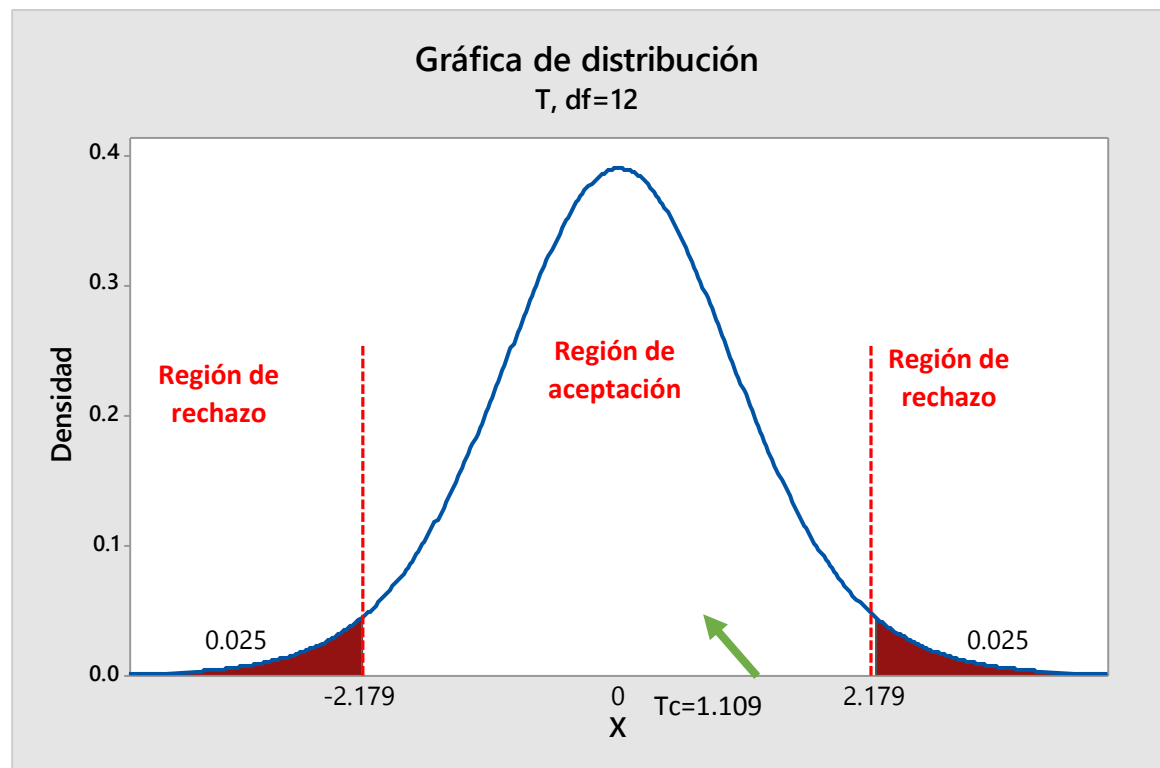


Figura 18. Gráfica de la hipótesis Cereales

Interpretación

Como $t=1.109 < t_c=2.179$ entonces aceptamos la hipótesis nula H_0 y nos quedamos con la hipótesis alterna H_a , en el nivel de significancia de $\alpha =5\%$. El cual afirma la hipótesis alterna H_a , que dice: Si se aplica el aspecto nutricional cereales apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Prueba de hipótesis complementos vitamínicos

Tabla 15. Correlación de la variable complementos vitamínicos

		complementos vitamínicos	Morbilidad Total por día
Complementos vitamínicos	Correlación de Pearson	1	,352
	Sig. (bilateral)		,217
	N	14	14
Morbilidad Total por día	Correlación de Pearson	,352	1
	Sig. (bilateral)	,217	
	N	14	14

Hipótesis Nula

H_0 = Si se aplica el aspecto nutricional complementos vitamínicos apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

H_a = Si se aplica el aspecto nutricional complementos vitamínicos apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D_1,y} = 0 \\ H_1 : \rho_{D_1,y} \neq 0 \end{cases}$$

$$t = \rho_{D_1,y} \sqrt{\frac{n-2}{1-\rho_{D_1,y}^2}}$$

$$t = 0.352 \sqrt{\frac{14-2}{1-(0.352)^2}}$$

$$t = 1.303$$

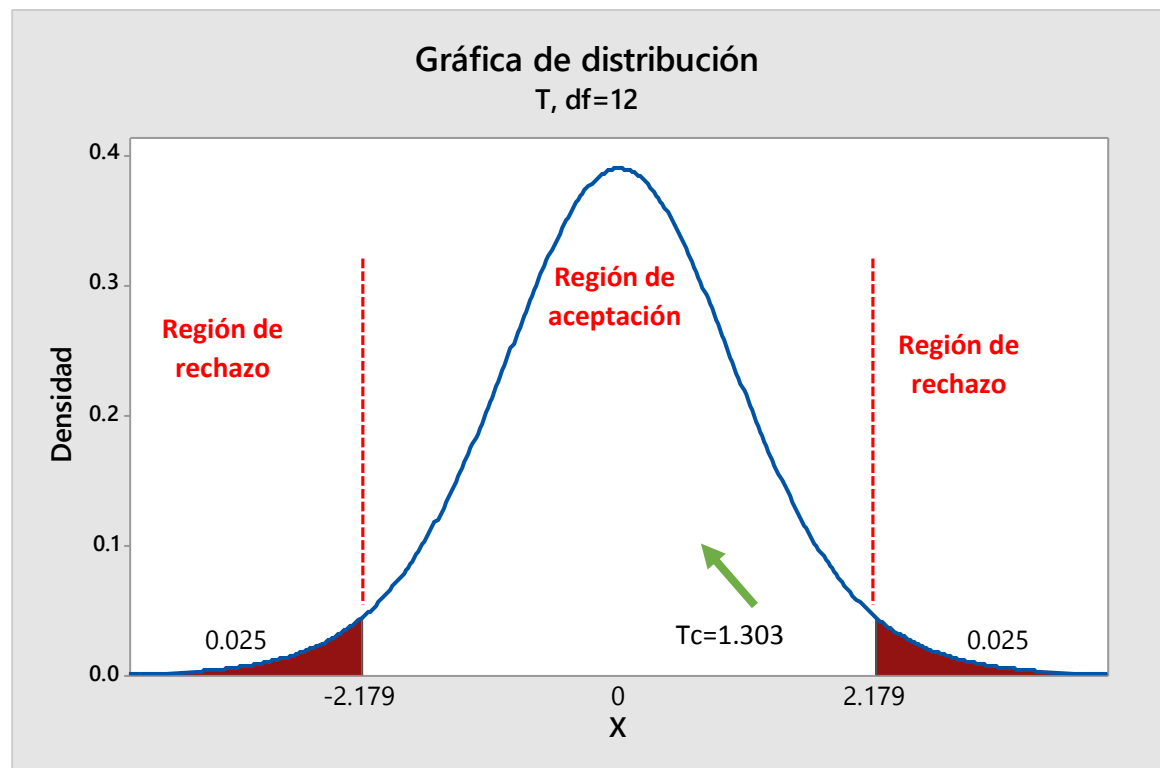


Figura 19. Gráfica de la hipótesis complementos vitamínicos

Interpretación

Como $t=1.303 < t_c=2.179$ entonces aceptamos la hipótesis nula H_0 , en el nivel de significancia de $\alpha =5\%$. El cual afirma la hipótesis nula H_0 . Que dice: Si se aplica el aspecto nutricional complementos vitamínicos apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Prueba de hipótesis condimentos

Tabla 16. Correlación de la variable condimentos

			condimentos	Morbilidad
Rho de Spearman	Condimentos	Coeficiente de correlación	1,000	,399
		Sig. (bilateral)	.	,158
		N	14	14
	Morbilidad	Coeficiente de correlación	,399	1,000
		Sig. (bilateral)	,158	.
		N	14	14

Hipótesis Nula

$H_0=$ Si se aplica el aspecto nutricional condimentos apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

$H_a=$ Si se aplica el aspecto nutricional condimentos apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D,y} = 0 \\ H_1 : \rho_{D,y} \neq 0 \end{cases}$$

$$t = \rho_{D,y} \sqrt{\frac{n-2}{1-\rho_{D,y}^2}}$$

$$t = 0.305 \sqrt{\frac{14-2}{1-(0.399)^2}}$$

$$t = 1.507$$

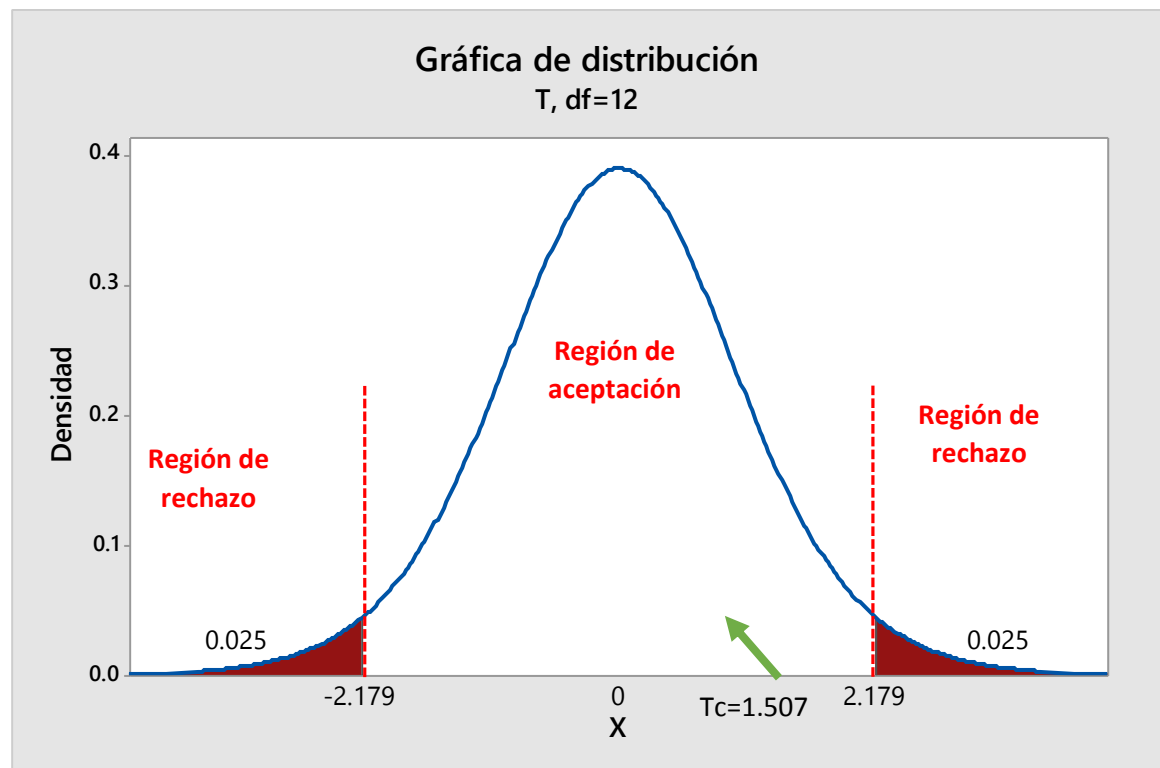


Figura20. Gráfica de la hipótesis condimentos

Interpretación

Como $t=1.507 < t_c=2.179$ entonces aceptamos la hipótesis nula H_0 y nos quedamos con la hipótesis alterna H_a , en el nivel de significancia de $\alpha =5\%$. El cual afirma la hipótesis alterna H_a . que dice: Si se aplica el aspecto nutricional condimentos apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Prueba de hipótesis Descalsificantes

Tabla 17. Correlación de la variable descalsificantes

			Descalcificante	Morbilidad
Rho de Spearman	Descalcificante	Coefficiente de correlación	1,000	,547*
		Sig. (bilateral)	.	,043
		N	14	14
Morbilidad	Morbilidad	Coefficiente de correlación	,547*	1,000
		Sig. (bilateral)	,043	.
		N	14	14

Hipótesis Nula

H_0 = Si se aplica el aspecto nutricional descalsificantes apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

H_a = Si se aplica el aspecto nutricional descalsificantes apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D,y} = 0 \\ H_1 : \rho_{D,y} \neq 0 \end{cases}$$

$$t = \rho_{D,y} \sqrt{\frac{n-2}{1-\rho_{D,y}^2}}$$

$$t = 0.547 \sqrt{\frac{14-2}{1-(0.547)^2}}$$

$$t = 2.264$$

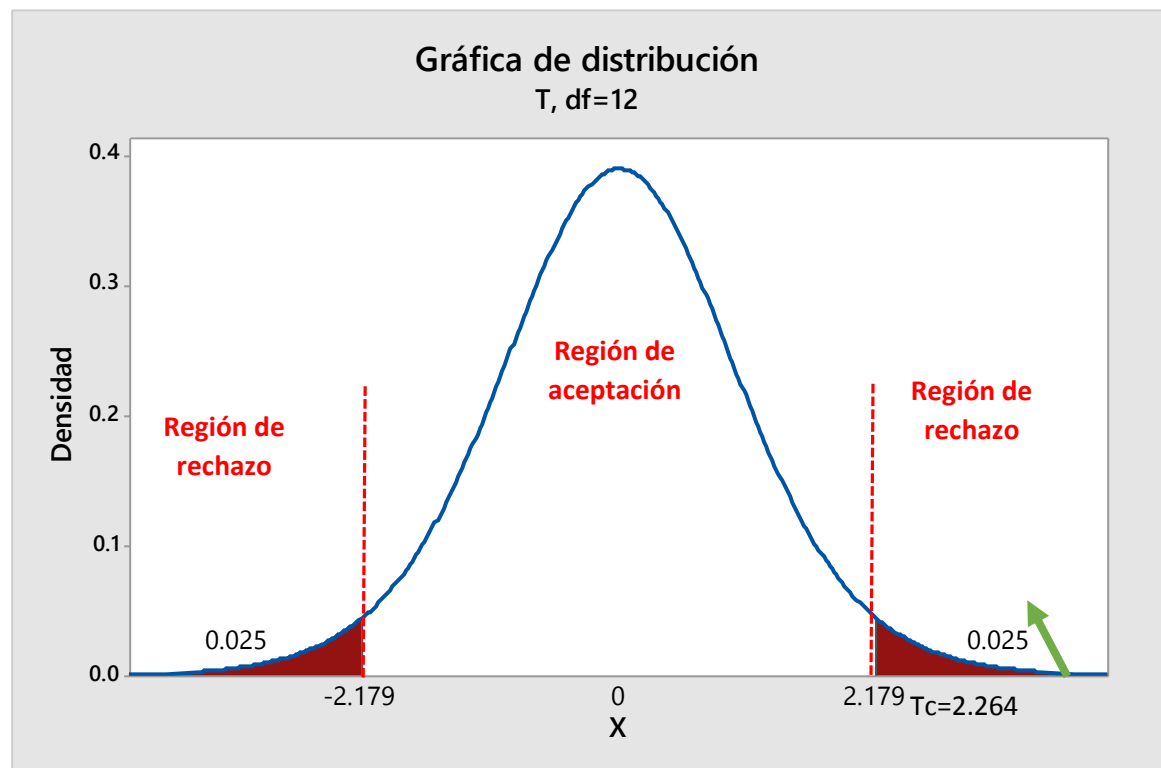


Figura 21. Gráfica de la hipótesis descalificantes

Interpretación

Como $t=2.264 > t_c=2.179$ entonces rechazamos la hipótesis nula H_0 y nos quedamos con la hipótesis alterna H_a , en el nivel de significancia de $\alpha =5\%$. El cual afirma la hipótesis alterna H_a que dice: Si se aplica el aspecto nutricional descalsificantes apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Prueba de hipótesis frutas

Tabla 18. Correlación de la variable frutas

			Frutas	Morbilidad
Rho de Spearman	Frutas	Coefficiente de correlación	1,000	,023
		Sig. (bilateral)	.	,938
		N	14	14
	Morbilidad	Coefficiente de correlación	,023	1,000
		Sig. (bilateral)	,938	.
		N	14	14

Hipótesis Nula

H_0 = Si se aplica el aspecto nutricional frutas apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

H_a = Si se aplica el aspecto nutricional frutas apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D,y} = 0 \\ H_1 : \rho_{D,y} \neq 0 \end{cases}$$

$$t = \rho_{D,y} \sqrt{\frac{n-2}{1-\rho_{D,y}^2}}$$

$$t = 0.023 \sqrt{\frac{14-2}{1-(0.023)^2}}$$

$$t = 0.080$$

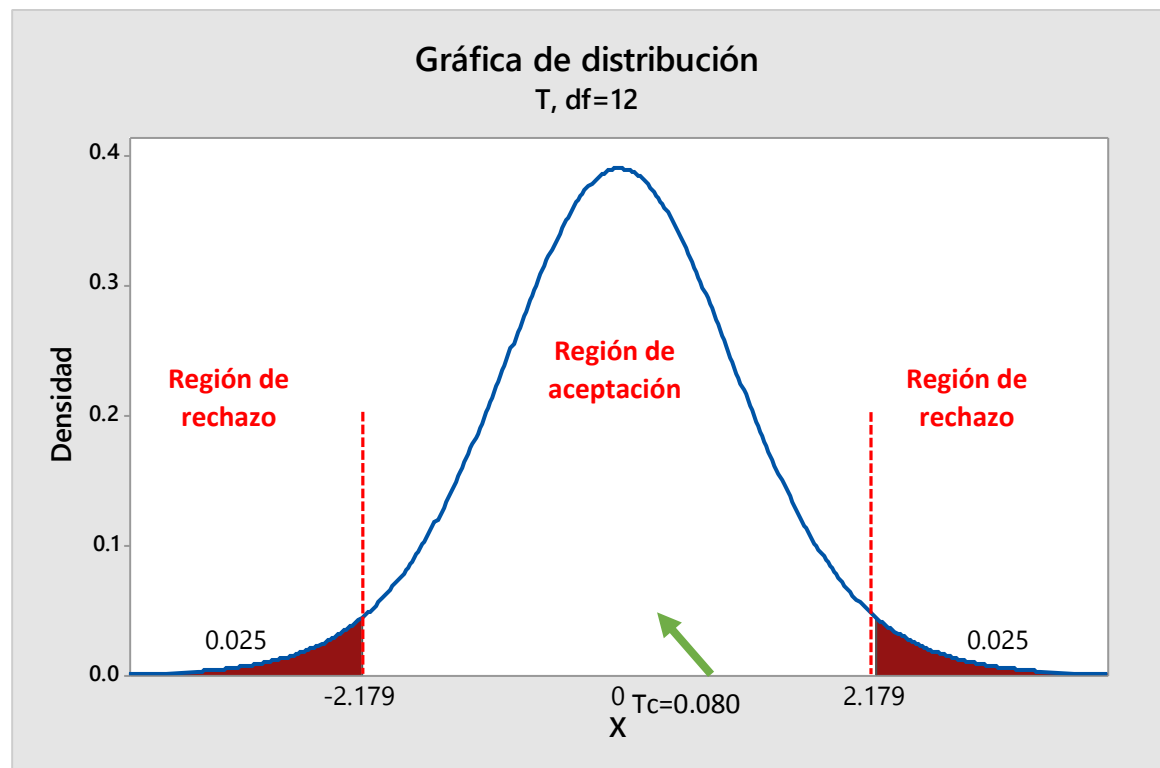


Figura22. Gráfica de la hipótesis frutas

Interpretación

Como $t=0.080 < t_c=2.179$ entonces aceptamos la hipótesis nula H_0 y nos quedamos con la hipótesis alterna H_a , en el nivel de significancia de $\alpha =5\%$. El cual afirma la hipótesis alterna H_a , que dice: Si se aplica el aspecto nutricional frutas apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Prueba de hipótesis panes

Tabla 19. Correlación de la variable panes

			Panes	Morbilidad
Rho de Spearman	Panes	Coeficiente de correlación	1,000	-,023
		Sig. (bilateral)	.	,939
		N	14	14
	Morbilidad	Coeficiente de correlación	-,023	1,000
		Sig. (bilateral)	,939	.
		N	14	14

Hipótesis Nula

$H_0=$ Si se aplica el aspecto nutricional panes apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

$H_a=$ Si se aplica el aspecto nutricional panes apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D,y} = 0 \\ H_1 : \rho_{D,y} \neq 0 \end{cases}$$

$$t = \rho_{D,y} \sqrt{\frac{n-2}{1-\rho_{D,y}^2}}$$

$$t = -0.023 \sqrt{\frac{14-2}{1-(-0.023)^2}}$$

$$t = -0.080$$

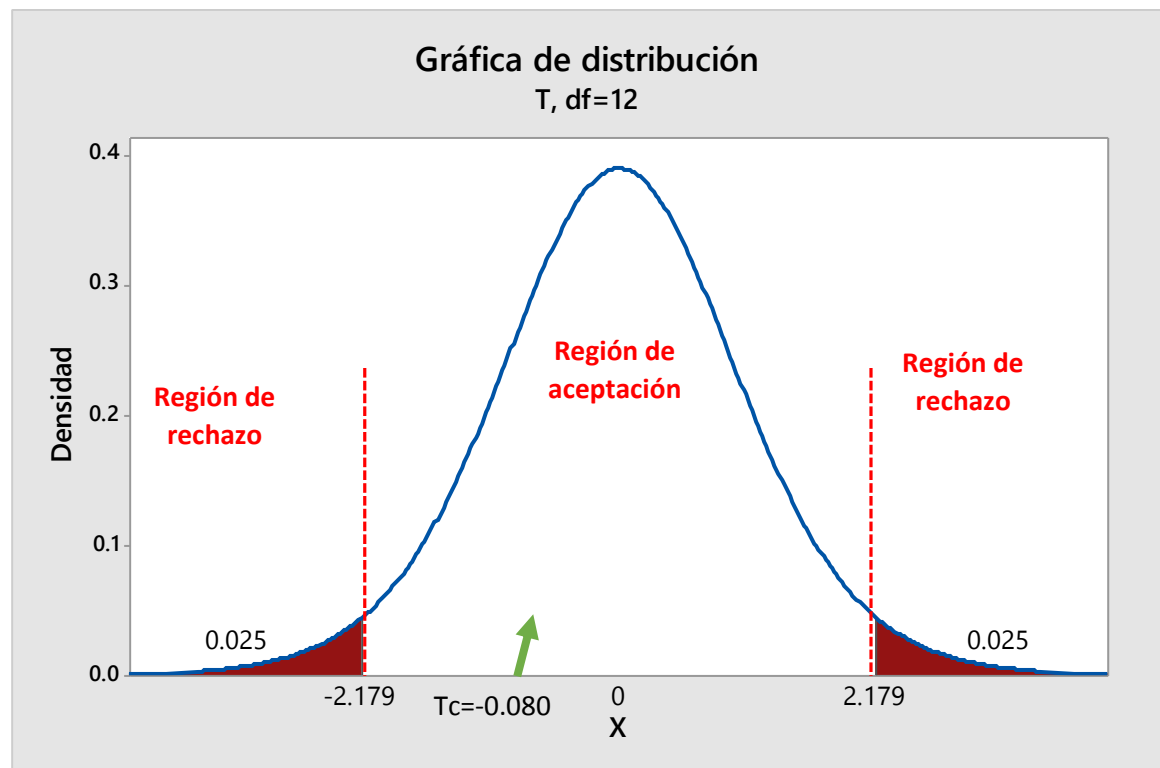


Figura 23. Gráfica de la hipótesis p-valor

Interpretación

Como $t = -0.080 < t_c = 2.179$ entonces aceptamos la hipótesis nula H_0 y nos quedamos con la hipótesis alterna H_a , en el nivel de significancia de $\alpha = 5\%$. El cual afirma la hipótesis alterna H_a , que dice: Si se aplica el aspecto nutricional panes apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Prueba de hipótesis productos lácteos

Tabla 20. Correlación de la variable productos lácteos

			Producto lácteos	Morbilidad
Rho de Spearman	Producto lácteos	Coeficiente de correlación	1,000	-,102
		Sig. (bilateral)	.	,728
		N	14	14
	Morbilidad	Coeficiente de correlación	-,102	1,000
		Sig. (bilateral)	,728	.
		N	14	14

Hipótesis Nula

$H_0 =$ Si se aplica el aspecto nutricional productos lácteos apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

$H_a =$ Si se aplica el aspecto nutricional productos lácteos apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D,y} = 0 \\ H_1 : \rho_{D,y} \neq 0 \end{cases}$$

$$t = \rho_{D,y} \sqrt{\frac{n-2}{1-\rho_{D,y}^2}}$$

$$t = -0.102 \sqrt{\frac{14-2}{1-(-0.102)^2}}$$

$$t = -0.355$$

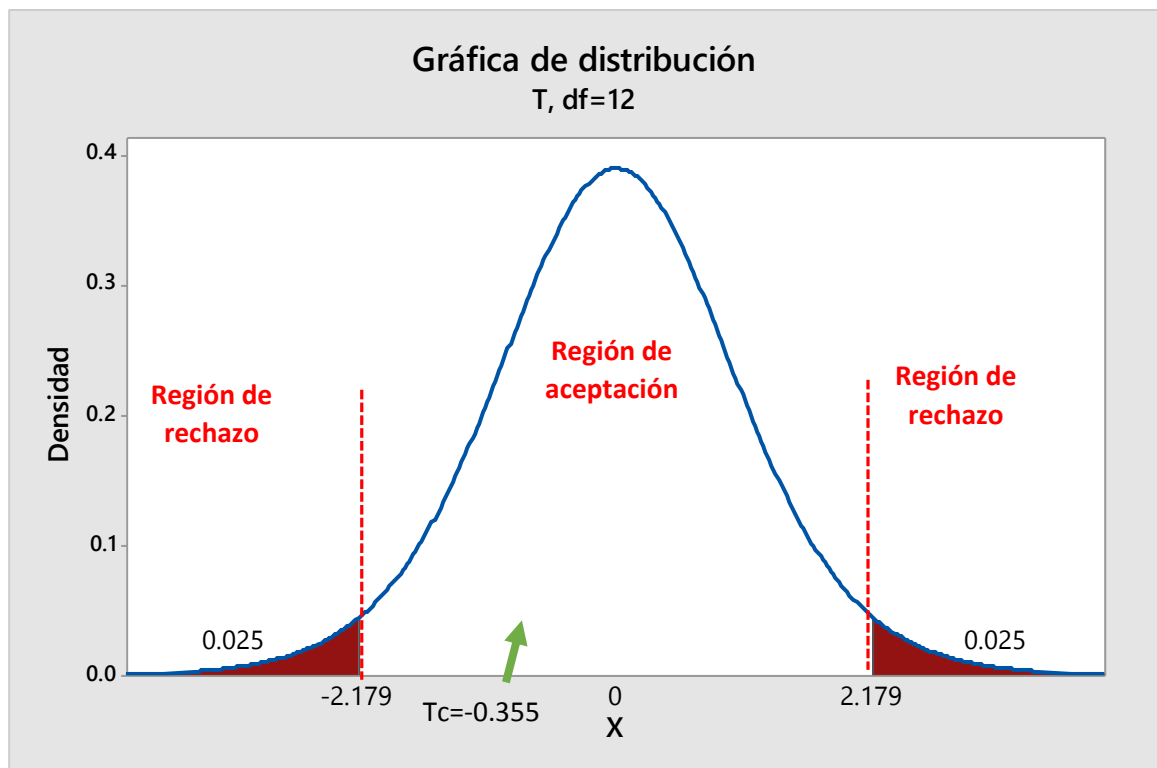


Figura 24. Gráfica de la hipótesis productos lácteos

Interpretación

Como $t = -0.355 < t_c = 2.179$ entonces aceptamos la hipótesis nula H_0 y nos quedamos con la hipótesis alterna H_a , en el nivel de significancia de $\alpha = 5\%$. El cual afirma la hipótesis alterna H_a , que dice: Si se aplica el aspecto nutricional productos lácteos apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Prueba de hipótesis sal

Tabla 21. Correlación de la variable sal

			Sal	Morbilidad
Rho de Spearman	Sal	Coeficiente de correlación	1,000	,641*
		Sig. (bilateral)	.	,013
		N	14	14
	Morbilidad	Coeficiente de correlación	,641*	1,000
		Sig. (bilateral)	,013	.
		N	14	14

Hipótesis Nula

$H_0 =$ Si se aplica el aspecto nutricional sal apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

$H_a =$ Si se aplica el aspecto nutricional sal apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D,y} = 0 \\ H_1 : \rho_{D,y} \neq 0 \end{cases}$$

$$t = \rho_{D,y} \sqrt{\frac{n-2}{1-\rho_{D,y}^2}}$$

$$t = 0.641 \sqrt{\frac{14-2}{1-(0.641)^2}}$$

$$t = 2.893$$

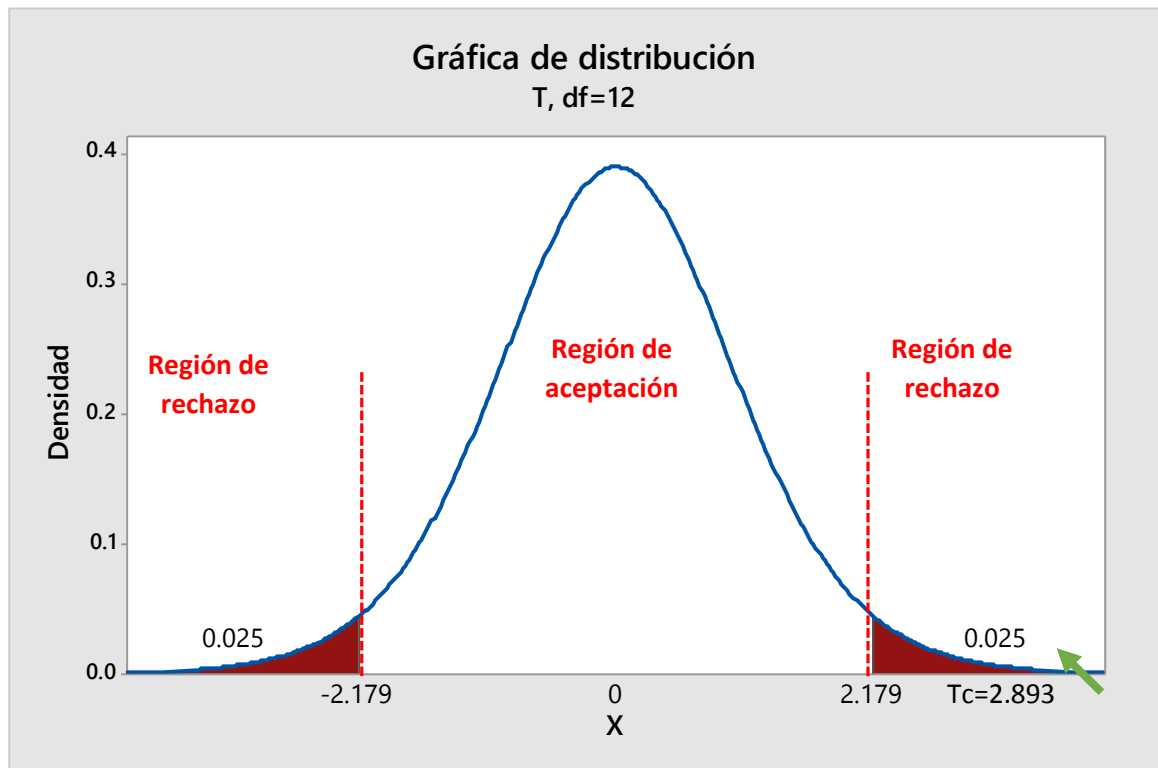


Figura25. Gráfica de la hipótesis sal

Interpretación

Como $t=2.893 > t_c=2.179$ entonces rechazamos la hipótesis nula H_0 y nos quedamos con la hipótesis alterna H_a , en el nivel de significancia de $\alpha =5\%$. El cual afirma la hipótesis alterna H_a , que dice: Si se aplica el aspecto nutricional sal apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Prueba de hipótesis verduras y hortalizas

Tabla 22. Correlación de la variable verduras y hortalizas

			Verduras y hortalizas	Morbilidad
Rho de Spearman	Verduras y hortalizas	Coefficiente de correlación	1,000	,551*
		Sig. (bilateral)	.	,041
		N	14	14
	Morbilidad	Coefficiente de correlación	,551*	1,000
		Sig. (bilateral)	,041	.
		N	14	14

*. La correlación es significativa en el nivel 0,05 (2 colas).

Hipótesis Nula

H_0 = Si se aplica el aspecto nutricional verduras y hortalizas apoyado por un Modelo Bayesiano entonces no existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Alterna

H_a = Si se aplica el aspecto nutricional verduras y hortalizas apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

Hipótesis Estadística

Puesto que los datos siguen una distribución **t** de **Estudent** y la muestra de investigación es $n=14$ que es menor a 30, entonces se aplica la prueba del estadístico t-student.

Sean:

$$\begin{cases} H_0 : \rho_{D,y} = 0 \\ H_1 : \rho_{D,y} \neq 0 \end{cases}$$

$$t = \rho_{D,y} \sqrt{\frac{n-2}{1-\rho_{D,y}^2}}$$

$$t = 0.551 \sqrt{\frac{14-2}{1-(0.551)^2}}$$

$$t = 2.287$$

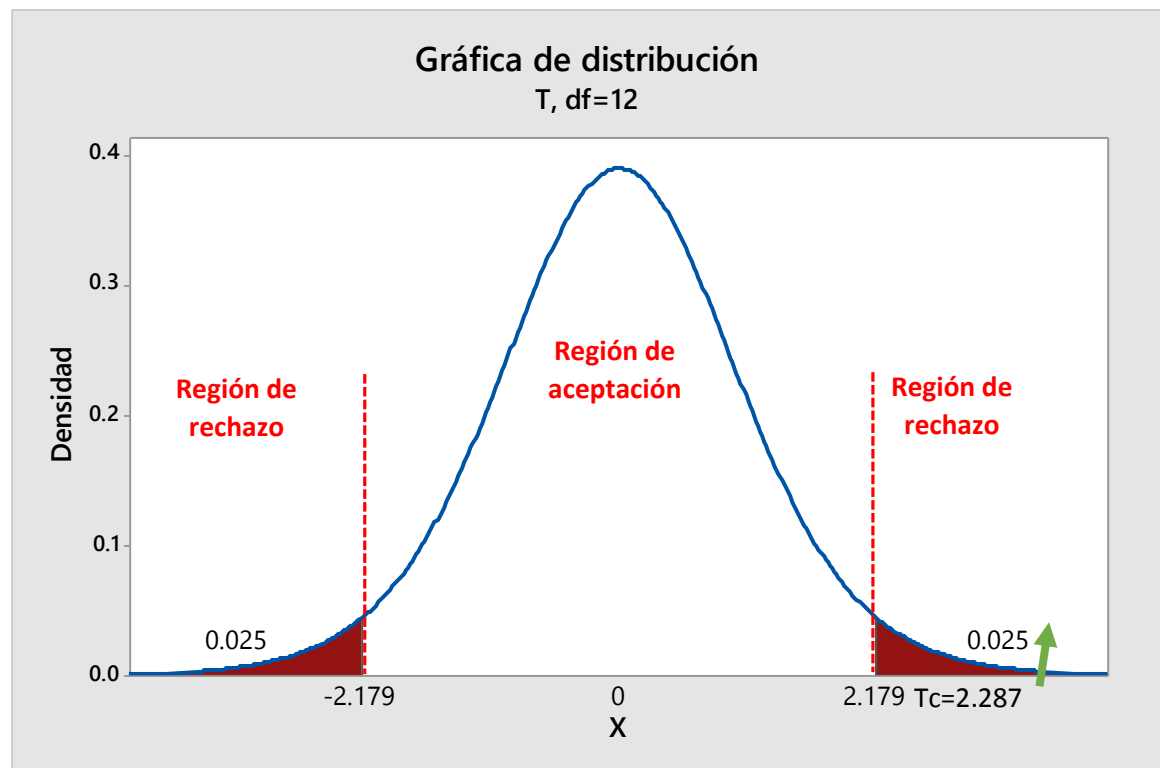


Figura 26. Gráfica de la hipótesis verduras y hortalizas

Interpretación

Como $t=2.287 > t_c=2.179$ entonces rechazamos la hipótesis nula H_0 y nos quedamos con la hipótesis alterna H_a , en el nivel de significancia de $\alpha =5\%$. El cual afirma la hipótesis alterna H_a que dice: Si se aplica el aspecto nutricional verduras y hortalizas apoyado por un Modelo Bayesiano entonces si existe correlación con el Pronóstico de la Condición de Morbilidad del Neonato de la gestante.

4.7 Descripción de los escenarios de pronóstico

En la investigación se trabajó con 4 escenarios, las cuales son las siguientes, según las escalas que se manejaron:

Alto mayor al 50% que el neonato salga con alguna deficiencia.

Medio 30% que el neonato salga con alguna enfermedad.

Bajo 1% que el neonato salga con enfermedad.

Mínimo 0 posibilidad que el neonato salga con enfermedad.

CAPÍTULO V:

RESULTADOS DE LA INVESTIGACIÓN

5.1 Resultados de los escenarios pronóstico

Las etapas para el desarrollo para la interfaz web, se detalla a continuación:

Etapas de Inicio

Se empezó a seleccionar las posibles arquitecturas que podrían usarse en el desarrollo del software, se identificaron los riesgos que se tendría durante el proyecto así como se estimaron algunos costos.

Tabla 23. Tabla de códigos de riesgo

CÓDIGO DE RIESGO	DESCRIPCIÓN
R01-2015	Falta de conocimiento sobre la implementación de la solución.
R02-2015	Falta de equipos para la elaboración del proyecto.
R03-2015	Mala planificación de los tiempos de construcción.
R04-2015	Ausencia de un miembro del equipo de desarrollo.
R05-2015	Falta de Software Licenciado o libre.

Tabla 24. Presupuestos de costos y recursos humanos en el desarrollo

Cargo	Costos S./
Jefe de Proyecto	3500
Analista	3000
Programador	2500
Especialista redes bayesianas	800

Asesor estadístico	800
Asesor de proyecto	800
Total	11400

Tabla 25. Presupuesto del hardware

Descripción	Costos S./
Pc	2500
Impresora multifuncional	800
Cámara fotográfica	800
Total	4100

Tabla 26. Presupuesto de software

Producto	Costo S./
Visual studio.net 2013	2167.45
Sqlserver 2012	2589.75
Excel 2013	398.65
Word 2013	398.65
Netica	1842.75
Total	7397.25

Etapas de Elaboración

Se seleccionó la arquitectura con la cual se trabajó el proyecto.

ARQUITECTURA N-CAPAS

Se seleccionó esta arquitectura por la facilidad para trabajar con un nivel superior de abstracción de los datos que se recopilan durante la ejecución del software.

También porque su bajo o casi nulo acoplamiento de paquetes de software lo cual brinda el beneficio de poder desarrollar un software escalable y de fácil mantenimiento.

Arquitectura N-Capas con Orientación al Dominio

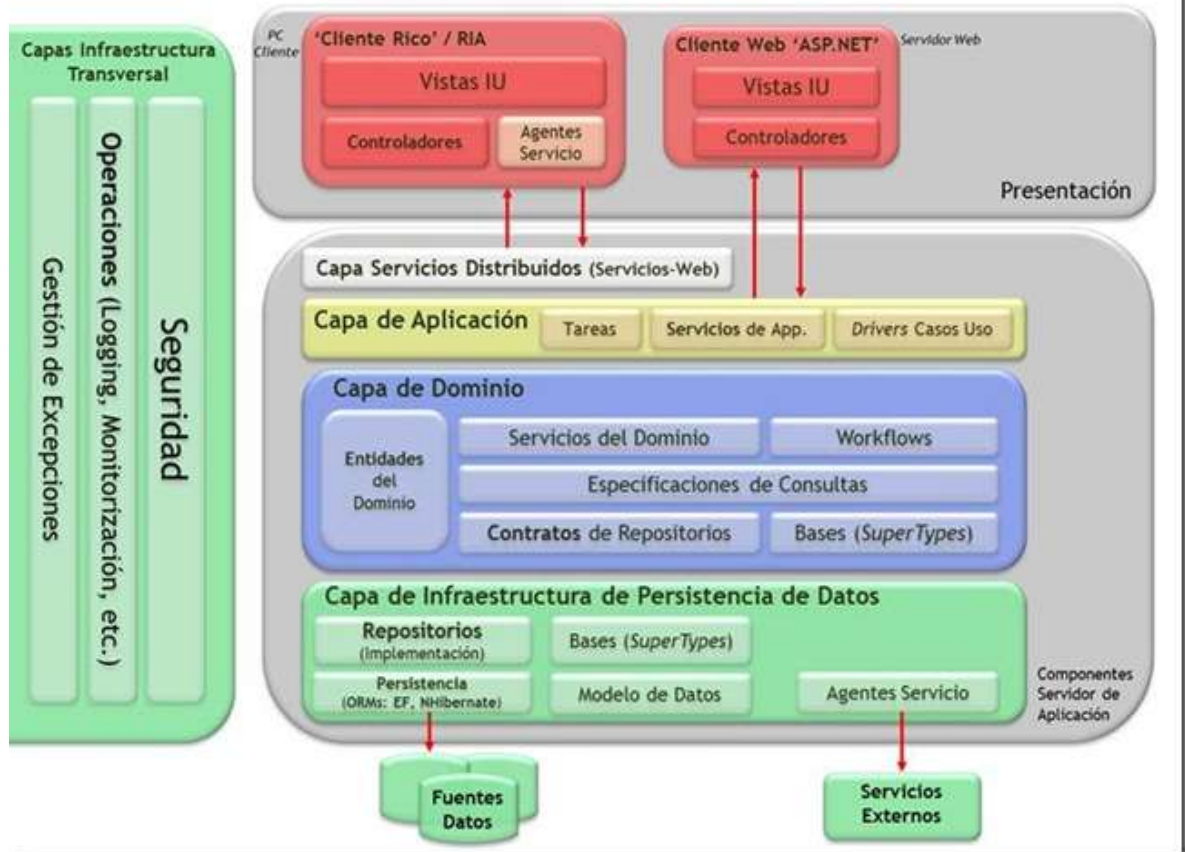


Figura27. Arquitectura n - capas

Las buenas prácticas para el desarrollo de software publicado por Microsoft, indica que es necesario trabajar en el desarrollo de un software que no esté altamente acoplado con otros paquetes, esto con el fin de que se pueda refactorizar, cambiar un paquete completo sí que esto afecto la performance de las mismas:

Etapa de Construcción

Se empezó a trabajar en el desarrollo del algoritmo, esto basado en el teorema de Bayes, el cual tiene la siguiente formula:

$$P(A|B) = \frac{P(A) \times P(B|A)}{\sum P(A) \times P(B|A)}$$

Para ello ya se definieron anteriormente las categorías a trabajar:

Las cuales quedaron clasificadas de la siguiente manera:

Nombre Dimensión
Cereales
Panes
Verduras y Hortalizas
Frutas
Aceites y grasas
Productos Lácteos
Carnes, Legumbres y Huevos
Complementos Vitamínicos
Agua
Descalcificantes
Sal
Azúcar
Condimentos

Figura 28. Dimensiones del sistemas

Cada una de las categorías antes mencionadas posee un valor máximo y un valor mínimo de consumo (desde, hasta), dichos valores fueron obtenidos mediante la agrupación de los alimentos de acuerdo a la pirámide alimenticia.

Se consideró la cantidad de calorías que aportaba cada grupo en base a una ración de 100g todos considerados en consumos diarios.

Se utiliza la información brindada en el libro de Dieta terapia de KRAUSS.

Los valores finales luego de la selección fueron los siguientes:

Nombre Dimensión	Desde	Hasta
Cereales	400	420
Panes	2	3
Verduras y Hortalizas	400	420
Frutas	4	5
Aceites y grasas	0	1
Productos Lácteos	400	420
Carnes, Legumbres y Huevos	200	300
Complementos Vitamínicos	1	2
Agua	7	8
Descalcificantes	0	0
Sal	0	4
Azúcar	0	24
Condimentos	0	4

Figura 29. Dimensiones y sus escalas

Estos valores son comparados con el total de las porciones consumidas por la madre gestante con la intención de medir su nivel de consumo.

ESCALA	DESCRIPCIÓN
NORMAL	Cuando el consumo Total diario esta DENTRO del los limites de consumo establecidos
BAJO	Cuando el consumo Total diario es MENOR al los limites de consumo establecidos
EXCESO	Cuando el consumo Total diario es SUPERIOR a los limites de consumo establecidos

Figura30. Escala de valores de las dimensiones

Para el registro de 14 días de raciones de alimentos ingeridos de una madre gestante se tendría las siguientes cantidades totales:

Nombre Dimensión	Desde	Hasta	x 14 días	x 14 días	Sumatoria de Raciones
Cereales	400	420	5600	5880	5880
Panes	2	3	28	42	25
Verduras y Hortalizas	400	420	5600	5880	8190
Frutas	4	5	56	70	59
Aceites y grasas	0	1	0	14	76
Productos Lácteos	400	420	5600	5880	2100
Carnes, Legumbres y Huevos	200	300	2800	4200	3300
Complementos Vitamínicos	1	2	14	28	11
Agua	7	8	98	112	90
Descalcificantes	0	0	0	0	700
Sal	0	4	0	56	116
Azúcar	0	24	0	336	312
Condimentos	0	4	0	56	68

Figura31. Raciones de alimentos ingeridos de una madre gestante

Se calcula la probabilidad de cada uno de las categorías:

Esto dividiendo el total de cada categoría entre la CANTIDAD TOTAL (20927) dando

el siguiente resultado:

Nombre Dimensión	Desde	Hasta	x 14 días	x 14 días	Sumatoria de Raciones	PROBABILIDAD
Cereales	400	420	5600	5880	5880	0.281
Panes	2	3	28	42	25	0.001
Verduras y Hortalizas	400	420	5600	5880	8190	0.391
Frutas	4	5	56	70	59	0.003
Aceites y grasas	0	1	0	14	76	0.004
Productos Lácteos	400	420	5600	5880	2100	0.100
Carnes, Legumbres y Huevos	200	300	2800	4200	3300	0.158
Complementos Vitamínicos	1	2	14	28	11	0.001
Agua	7	8	98	112	90	0.004
Descalcificantes	0	0	0	0	700	0.033
Sal	0	4	0	56	116	0.006
Azúcar	0	24	0	336	312	0.015
Condimentos	0	4	0	56	68	0.003
				TOTAL	20927	1

Figura 32. Cálculo de la probabilidad

Una vez obtenida la probabilidad de cada una de las categorías, se evalúa cuáles fueron las categorías en las cuales la suma total de sus raciones sobrepasa el rango permitido.

Cada una de las categorías tienen un valor de cuanto afecta a la morbilidad en caso exista un consumo bajo o exceso.

Los mismos se detallan en el siguiente cuadro:

Nombre Dimensión	Bajo	Exceso
Cereales	0.55	0.25
Panes	0.02	1
Verduras y Hortalizas	0.55	0.25
Frutas	0.55	0.25
Aceites y grasas	0.01	1
Productos Lácteos	0.6	0.3
Carnes, Legumbres y Huevos	0.5	1
Complementos Vitamínicos	0.4	1
Agua	0.7	0.7
Descalcificantes	0	1
Sal	0	1
Azúcar	0.02	1
Condimentos	0	1

Figura 33. Valor de categoría, según las dimensiones

Si el total de la ración es normal la categoría toma un valor 0, cuando es bajo exceso toma el valor indicado en la columna bajo estimando el porcentaje que representa el mismo sobre el rango permitido de ración.

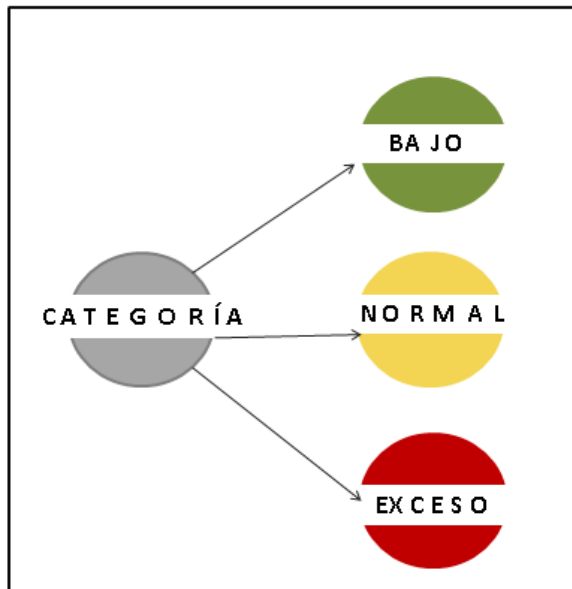


Figura 34. Categoría y sus escalas

Ejemplo:

Si una madre tomó durante el día 7 vasos de agua pero debía tomar como mínimo 8 vasos, entonces le faltó 1 vaso para completar lo requerido diariamente, entonces tuvo un consumo BAJO de agua.

Entonces se evalúa cuanto es el porcentaje que este representa, pues no es lo mismo que le falte 1 vaso de agua para completar lo requerido a que te falte 5 vasos.

El cálculo se realiza del siguiente modo:

CONSUMO BAJO	
Vasos de Agua consumidos:	7
Requeridos	8
faltante	1
Porcentaje faltante:	12.50%
Nombre Dimensión	Bajo
Agua	0.7
VALOR FINAL (12.5% de 0.70)	0.09

Figura 35. Cálculo de porcentaje de vasos por día

Finalmente, una vez obtenido todos los valores según las categorías y los días (14) se tiene lo siguiente:

Nombre Dimensión	Desde	Hasta	x 14 días	x 14 días	Sumatoria de Raciones	PROBABILIDAD	CALCULADO
Cereales	400	420	5600	5800	5800	0.291	0
Panes	2	3	28	42	25	0.001	0
Verduras y Hortalizas	400	420	5600	5800	8190	0.291	0.07
Frutas	4	5	56	70	59	0.003	0
Aceites y grasas	0	1	0	14	76	0.004	0.02
Productos Lácteos	400	420	5600	5800	2100	0.100	0.38
Carnes, Legumbres y Huevos	200	300	2800	4200	3300	0.158	0
Complementos Vitamínicos	1	2	14	28	11	0.001	0.24
Agua	7	8	98	112	90	0.004	0.06
Descalcificantes	0	0	0	0	700	0.033	1
Sal	0	4	0	56	116	0.006	0.52
Azúcar	0	24	0	336	312	0.015	0
Condimentos	0	4	0	56	68	0.003	0.18
				TOTAL	20927	1	

Figura 36. Valores por día

Entonces se aplica la fórmula de Bayes (Probabilidad, calculado):

$$P(A|B) = \frac{P(A) \times P(B|A)}{\sum P(A) \times P(B|A)}$$

Considerando que al resultado final se le debe agregar el riesgo en cuanto a la edad de la madre considerando que para las pruebas se recogió la información de una gestante de 34 años y el Índice de masa corporal.

Tenemos un valor final de: 0.46 que representa el 46%

Según lo explicado anteriormente se realizó la codificación del algoritmo:

Selección de la información Mediante consulta SQL:

```

25
26 select @dias GATEDIFF(day,MIN(d.fecha),MAX(d.fecha))-1 from Dieta d
27 where
28 MONTH(d.fecha)= MONTH(GETDATE()) and
29 YEAR(d.fecha)=YEAR(GETDATE()) and
30 d.Idusuario= @Idusuario
31
32 Insert into @Table (NombreDimension,Idcategoria,Medida,racion,totalracion,porcionNormaldesde,
33 porcionNormalhasta,condicion,probabilidad)
34 select
35 c.NombreDimension,
36 c.Idcategoria,
37 n.Medida,
38 sum(dd.racion),
39 sum(dd.racion)*n.Medida as totalracion,
40 dc.porcionNormaldesde*@dias as porcionNormaldesde,
41 dc.porcionNormalhasta*@dias as porcionNormalhasta,
42 CASE
43     WHEN ((sum(dd.racion)*n.Medida)>{dc.porcionNormaldesde*@dias}) and
44          ((sum(dd.racion)*n.Medida)>{dc.porcionNormalhasta*@dias})
45     THEN
46         'Exceso'
47     WHEN ((sum(dd.racion)*n.Medida)<{dc.porcionNormaldesde*@dias}) and
48          ((sum(dd.racion)*n.Medida)<{dc.porcionNormalhasta*@dias})
49     THEN
50         'Bajo'
51     ELSE 'Normal'
52 END condicion ,
53 CASE
54     WHEN ((sum(dd.racion)*n.Medida)>{dc.porcionNormaldesde*@dias}) and
55          ((sum(dd.racion)*n.Medida)>{dc.porcionNormalhasta*@dias})
56     THEN
57         (select ValorMaximo from CategoriaAlimento where Idcategoria=c.Idcategoria)
58     WHEN ((sum(dd.racion)*n.Medida)<{dc.porcionNormaldesde*@dias}) and
59          ((sum(dd.racion)*n.Medida)<{dc.porcionNormalhasta*@dias})
60     THEN
61         (select ValorMinimo from CategoriaAlimento where Idcategoria=c.Idcategoria)
62     ELSE 0
63 END probabilidad
64 from DetalleCategoria dc
65 Inner join CategoriaAlimento c on dc.Idcategoria=c.Idcategoria
66 Inner join Medidas m on dc.Idmedidam,Idmedida
67 Inner join Usuario u on dc.Idusuario=u.Idusuario
68 Inner join Persona p on u.Idperson=p.Idperson
69 Inner join DetalleDieta dd on dc.Idcategoria=dd.Idcategoria
70 Inner join Dieta d on dd.Iddieta=d.Iddieta
71 where
72 dc.Idusuario=@Idusuario and
73 MONTH(dc.fecha)= MONTH(GETDATE()) and
74 YEAR(dc.fecha)=YEAR(GETDATE())and
75 MONTH(d.fecha)= MONTH(GETDATE()) and
76 YEAR(d.fecha)=YEAR(GETDATE())
77

```

Figura 37. Codificación del algoritmo

Realizando llamado a la consulta de selección:

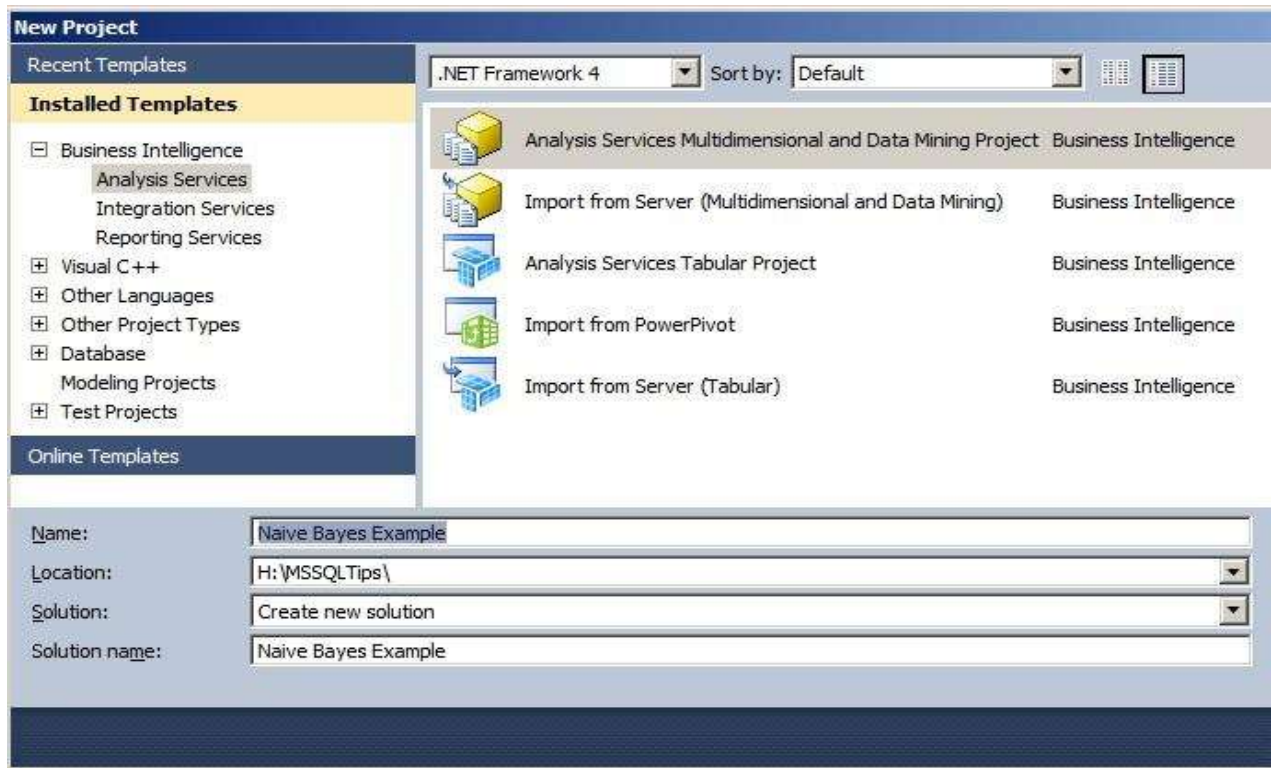


Figura 38. Uso del analysis services

Algoritmo de clasificación

```
1 namespace nBayes
2 {
3     using System;
4
5     public class Analyzer
6     {
7         private float I = 0;
8         private float invI = 0;
9
10        public Analyzer()
11        {
12            this.Tolerance = .05f;
13        }
14
15        public float Tolerance { get; set; }
16
17        public CategorizationResult Categorize(Entry item, Index first, Index second)
18        {
19            float prediction = GetPrediction(item, first, second);
20
21            if (prediction <= .5f - this.Tolerance)
22                return CategorizationResult.Second;
23
24            if (prediction >= .5 + this.Tolerance)
25                return CategorizationResult.First;
26
27            return CategorizationResult.Undetermined;
28        }
29
30        public float GetPrediction(Entry item, Index first, Index second)
31        {
32            foreach (string token in item)
33            {
34                int firstCount = first.GetTokenCount(token);
35                int secondCount = second.GetTokenCount(token);
36
37                float probability = CalcProbability(firstCount, first.EntryCount, secondCount, second.EntryCount);
38
39            }
40
41            float prediction = CombineProbability();
42            return prediction;
43        }
44
45        private float CalcProbability(float cat1count, float cat1total, float cat2count, float cat2total)
46        {
47            float bw = cat1count / cat1total;
48            float gw = cat2count / cat2total;
49            float pw = ((bw) / ((bw) + (gw)));
50            float
51                s = 1f;
52
53            n = cat1count + cat2count;
54            float fw = ((s * w) + (n * pw)) / (s + n);
55            LogProbability(fw);
56
57            return fw;
58        }
59
60        private void LogProbability(float prob)
61        {
62            if (!float.IsNaN(prob))
63            {
64                I = I == 0 ? prob : I * prob;
65                invI = invI == 0 ? (1 - prob) : invI * (1 - prob);
66            }
67        }
68
69        private float CombineProbability()
70        {
71            return I / (I + invI);
72        }
73    }
74 }
75 }
76 }
```

Figura 39. Codificación del algoritmo de clasificación

Entrenamiento del Modelo Bayesiano, utilizando la Tecnología de la Información.

Resultados usando el software Netica por dimensión:

El diagrama de arborización del Netica se generó en función a las 13 dimensiones del aspecto nutricional de las madres gestantes, el cual se identificaron 13 eventos principales que nos permitirá construir el Modelo Bayesiano propuesto por nuestra investigación. Luego generamos el árbol de decisión por las 13 dimensiones del aspecto nutricional, para luego realizar entrenamientos del modelo que nos permitan pronosticar la morbilidad del neonato. El entrenamiento del Modelo Bayesiano, se realizó con los datos según las dimensiones, escalas y los valores que se muestran en las figuras 28, 29 y 30, que fueron registrados en la interfaz web, en la que dichos datos alimentaron a cada uno de las dimensiones.

El proceso de entrenamiento se realizó según los escenarios que se muestran a continuación:

Escenarios 1. Definición de las variables declaradas y sus escalas, dentro del software Netica.

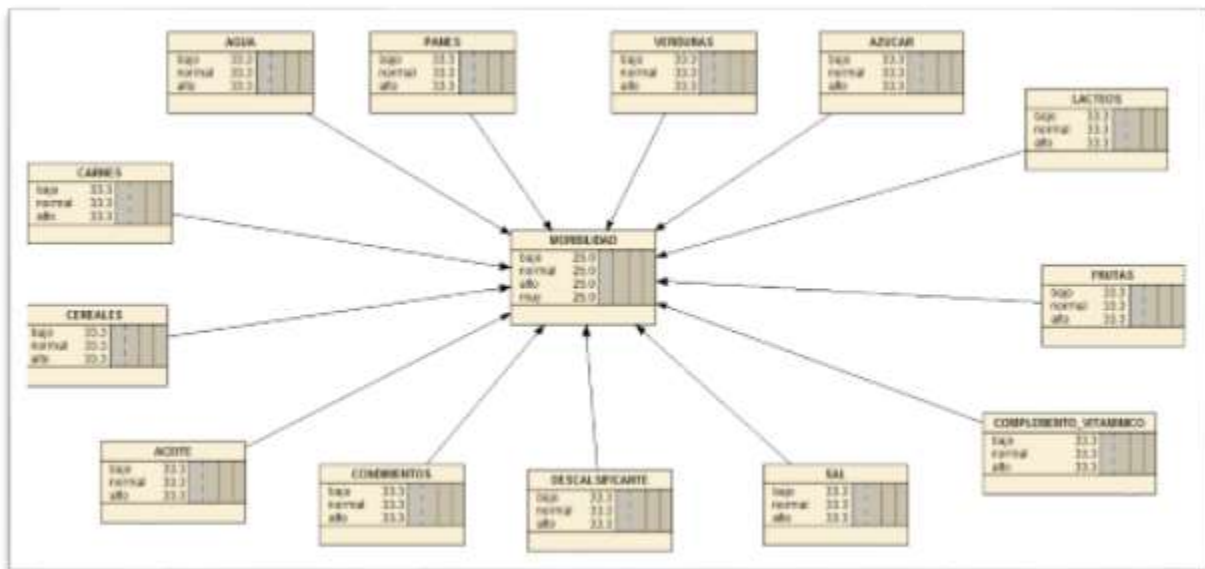


Figura 40. Definición de variables

Escenario 2. Ingreso de datos de la madre gestante (14 días), para entrenar la red neuronal.

modelo_1_modificado: Bloc de notas														
Archivo	Edición	Formato	Ver	Ayuda										
IMBILIDAD	ACEITE	AGUA	AZUCAR	CARNES	CEREALES	COMPLEMENTO_VITAMINICO	CONDIMENTOS	DESCALCIFICANTE	FRUTAS	PANES	LACTEOS	SAL	VERDURAS	
bajo	bajo	normal	normal	normal	bajo	bajo	normal	bajo	bajo	alto	normal	normal	bajo	
bajo	bajo	bajo	bajo	bajo	alto	bajo	bajo	normal	bajo	bajo	bajo	normal	alto	
bajo	bajo	normal	bajo	bajo	bajo	bajo	normal	bajo	bajo	bajo	bajo	normal	normal	
medio	normal	bajo	bajo	bajo	normal	bajo	bajo	bajo	bajo	bajo	bajo	alto	alto	
bajo	bajo	bajo	bajo	normal	normal	normal	bajo	bajo	bajo	bajo	bajo	alto	normal	
medio	alto	alto	alto	bajo	alto	bajo	normal	bajo	alto	normal	bajo	alto	alto	
medio	alto	bajo	alto	normal	normal	normal	alto	normal	alto	normal	normal	alto	alto	
medio	alto	alto	normal	ajo	alto	alto	alto	alto	bajo	normal	bajo	alto	alto	
alto	alto	bajo	bajo	alto	bajo	normal	alto	normal	bajo	normal	bajo	alto	normal	
bajo	alto	bajo	bajo	normal	bajo	alto	normal	bajo	bajo	normal	normal	alto	normal	
medio	alto	bajo	normal	alto	alto	normal	alto	bajo	normal	alto	bajo	alto	alto	
medio	alto	bajo	normal	bajo	alto	bajo	normal	bajo	alto	normal	bajo	alto	alto	
medio	normal	bajo	alto	bajo	bajo	alto	normal	normal	bajo	normal	bajo	alto	alto	
bajo	alto	normal	normal	bajo	bajo	normal	normal	bajo	alto	alto	bajo	normal	normal	

Figura 41. Ingreso de datos - Madre gestante

Escenario 3. Entrenamiento del Modelo Bayesiano con los datos de la madre gestante.

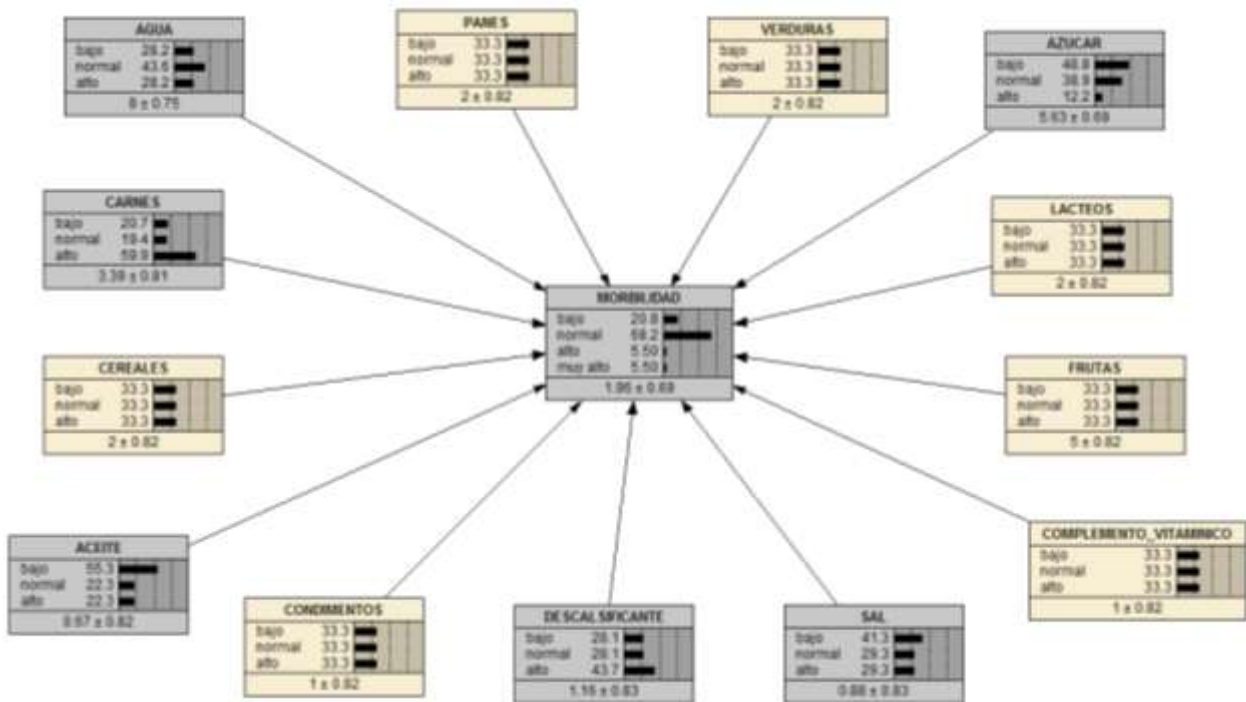


Figura 42. Entrenamiento del Modelo Bayesiano

CONCLUSIONES

Luego de haber cumplido los objetivos de la investigación, llegamos a las siguientes conclusiones específicas.

Específica 1. Al aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión aceites y grasas arrojó una correlación muy significativa de Spearman, cuyo valor es de 0.679 y así mismo observamos un sig bilateral igual 0.008, que por ser menor al nivel de significancia alfa del 5 % hace que afirmemos: “Si se aplica el aspecto nutricional aceites y grasas apoyado por un Modelo Bayesiano entonces si existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”.

Específica 2. Aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión azúcar arrojó una correlación de Spearman, cuyo valor es de 0.070 y así mismo observamos un sig bilateral igual 0.813, que por ser mayor al nivel de significancia alfa del 5 % afirmamos: “Si se aplica el aspecto nutricional azúcar apoyado por un Modelo Bayesiano entonces no existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”. Sin embargo, el valor de Rho de Spearman para esta dimensión no dice que la correlación es muy débil.

Específica 3. Al aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que la dimensión agua arrojó una correlación negativa de Spearman, cuyo valor es de -0.394 y así mismo observamos un sig bilateral igual 0.163, que por ser menor al nivel de significancia alfa del 5 % hace que afirmemos: “Si se aplica el aspecto nutricional agua apoyado por un Modelo Bayesiano

entonces si existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”.

Específica 4. Al aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión carnes y legumbres arrojó una correlación de Spearman, cuyo valor es de 0.175 y así mismo observamos un sig bilateral igual 0.549, que por ser mayor al nivel de significancia alfa del 5 % afirmamos: “Si se aplica el aspecto nutricional carnes y legumbres apoyado por un Modelo Bayesiano entonces no existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”. Sin embargo, el valor de Rho de Spearman para esta dimensión no dice que la correlación es muy débil.

Específica 5. Al aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión cereales arrojó una correlación muy significativa de Spearman, cuyo valor es de 0.305 y así mismo observamos un sig bilateral igual 0.289, que por ser menor al nivel de significancia alfa del 5 % hace que afirmemos: “Si se aplica el aspecto nutricional cereales apoyado por un Modelo Bayesiano entonces si existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”.

Específica 6. Al aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión complementos vitamínicos arrojó una correlación muy significativa de Spearman, cuyo valor es de 0.352 y así mismo observamos un sig bilateral igual 0.217, que por ser menor al nivel de significancia alfa del 5 % hace que afirmemos: “Si se aplica el aspecto nutricional cereales apoyado por un Modelo Bayesiano entonces si existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”.

Específica 7. Al aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión condimentos arrojó una correlación muy significativa de Spearman, cuyo valor es de 0.399 y así mismo observamos un sig bilateral igual 0.158, que por ser menor al nivel de significancia alfa del 5 % hace que afirmemos: “Si se aplica el aspecto nutricional condimentos apoyado por un Modelo Bayesiano entonces si existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”.

Específica 8. Al aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión descalsificantes arrojó una correlación muy significativa de Spearman, cuyo valor es de 0.547 y así mismo observamos un sig bilateral igual 0.043, que por ser menor al nivel de significancia alfa del 5 % hace que afirmemos: “Si se aplica el aspecto nutricional descalsificantesapoyado por un Modelo Bayesiano entonces si existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”.

Específica 9. Aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión frutas arrojó una correlación de Spearman, cuyo valor es de 0.023 y así mismo observamos un sig bilateral igual 0.938, que por ser mayor al nivel de significancia alfa del 5 % afirmamos: “Si se aplica el aspecto nutricional frutas apoyado por un Modelo Bayesiano entonces no existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”. Sin embargo, el valor de Rho de Spearman para esta dimensión no dice que la correlación es muy débil.

Específica 10. Aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión panes arrojó una

correlación de Spearman, cuyo valor es de 0.939 y así mismo observamos un sig bilateral igual 0.813, que por ser mayor al nivel de significancia alfa del 5 % afirmamos: “Si se aplica el aspecto nutricional panes apoyado por un Modelo Bayesiano entonces no existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”. Sin embargo, el valor de Rho de Spearman para esta dimensión no dice que la correlación es muy débil.

Especifica 11. Aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión productos lácteos arrojó una correlación de Spearman, cuyo valor es de -0.102 y así mismo observamos un sig bilateral igual 0.728, que por ser mayor al nivel de significancia alfa del 5 % afirmamos: “Si se aplica el aspecto nutricional azúcar apoyado por un Modelo Bayesiano entonces no existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”. Sin embargo, el valor de Rho de Spearman para esta dimensión no dice que la correlación es muy débil.

Especifica 12. Al aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión sal arrojó una correlación muy significativa de Spearman, cuyo valor es de 0.641 y así mismo observamos un sig bilateral igual 0.013, que por ser menor al nivel de significancia alfa del 5 % hace que afirmemos: “Si se aplica el aspecto nutricional sal apoyado por un Modelo Bayesiano entonces si existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”.

Especifica 13. Al aplicar la interfaz web a las madres gestantes para medir su morbilidad del neonato apoyado en el Modelo Bayesiano, se obtuvo que dimensión verduras y hortalizas arrojó una correlación muy significativa de Spearman, cuyo valor es de 0.551 y así mismo

observamos un sig bilateral igual 0.041, que por ser menor al nivel de significancia alfa del 5 % hace que afirmemos: “Si se aplica el aspecto nutricional verduras y hortalizas apoyado por un Modelo Bayesiano entonces si existe correlación con el pronóstico de la condición de morbilidad del neonato de la gestante”.

RECOMENDACIONES

Recomendación 1. En esta investigación, con el uso del modelo bayesiano, se recomienda que al momento de recoger la información se tenga en cuenta el número de días de registro por madre gestante para mejorar la variabilidad en los datos a fin de tomar buenas decisiones.

Recomendación 2. La investigación se debe realizar en el área de neonatología de hospitales. Las madres gestantes deben estar registradas para poder llevar un buen embarazo, ya que se encontró que no todas ellas tenían un teléfono móvil (Smartphone) lo que dificultaba el recojo de información en tiempo real. Para que ellas mismas verifiquen su proceso.

Recomendación 3. Para futuras investigaciones se sugiere usar el algoritmo K2 del modelo bayesiano, cuyo entrenamiento agilice los patrones de las variables en estudio con el objetivo de predecir eficazmente la morbilidad de las madres gestantes.

Recomendación 4. Para el uso de la interfaz web, se recomienda hacer una implementación de las mismas, hacia las madres gestantes, para que el control de la calidad de los registros, sean más efectivos.

REFERENCIAS BIBLIOGRÁFICAS

1. Antonio Gil Almira, 2008 “Embarazo: diagnóstico, edad gestacional y fecha del parto”.
2. Mortalidad Neonatal en el Perú y sus departamentos 2011 – 2012
http://www.dge.gob.pe/portal/index.php?option=com_content&view=article&id=440&Itemid=204.
3. Pedro Larrañaga, Iñaki Inza, Abdelmalik Moujahid, Departamento de Ciencias de la Computación e Inteligencia Artificial Universidad del País.
4. G. Rodriguez, V. Erraez, J. Mejia, Aprobación de un módulo utilizando probabilidades, a través de redes bayesianas, 2014.
5. Jorge Lopez Puga, Revista Electrónica de Metodología Aplicada 2012, Vol. 17 nº 1, pp. 1-17.
6. Gabriela Esperón, Alberto E. Laudadio, Alejandro Fiel Martíne, Eduardo Serrano, Redes neuronales y series de Fourier aplicados al procesamiento de imágenes en problemas de aprendizaje, Universidad de palermo, Facultad de ingeniería, Buenos Aires, Argentina.
7. Maria Isabel Angeles larrieta y Angélica Maria Santillán Gómez, Minería de datos: concepto, características, estructura y aplicaciones.
8. population Reports, "Espacio de los nacimientos con intervalos de tres a cinco años, más vidas a salvo", 2002.
9. Felipe Flores Perez, "Embriología humana".
10. Hübner G, María Eugenia Juárez H, "Test de Apgar. Después de medio siglo ¿sigue vigente?", 2002.
11. Oscar Munares García, Guillermo Gómez Guizado, Juan Barboza el Carpio, José Sánchez-Abanto, "Niveles de hemoglobina en gestantes atendidas en establecimientos del Ministerio de Salud del Perú, 2011", 2012.
12. V. Cararach Ramoneda y F. Botet Mussons, "preclamsia. Eclampsia y síndrome de Hellp", 2008.
13. Felipe Perez, "Periodo fetal", 2008.
14. kathleen mahan, silvia scoot. (2008). Krause Dietoterapia. España: masson.

15. Directrices y procedimientos recomendados de .NET Framework.
[https://msdn.microsoft.com/es-es/en-s/library/vstudio/ms184412\(v=vs.100\).aspx](https://msdn.microsoft.com/es-es/en-s/library/vstudio/ms184412(v=vs.100).aspx).
16. Jorge Sisalima, “Diagnóstico de enfermedades utilizando Redes Bayesianas”, 2010.
17. V. Messina, M. Cuevas, “Redes Bayesianas para predecir riesgo de plomo en sangre de puérperas y neonatos”, 2010.
18. Rocío Martínez, Patricia León, Nicandro Ramírez, Héctor Mesa, “Evaluación del Potencial de Redes Bayesianas en la Clasificación en Datos Médicos”, 2008.
19. José Hernandez Oralla, José Ramírez Quintana, Cesar Ferri Ramírez. (2004).
Introducción a la minería de datos. Barcelona - España: Editorial Alhambra S. A.
pgs. 258 – 278.
20. Sylvia Escott -Stump, L. Kathleen Mahan. (2008). Krause Dietoterapia 12. a edición. Barcelona - España: Elsevier Masson. Pgs 160 – 192.

ANEXOS

Anexo 1. Proceso de morbilidad materna y neonatal

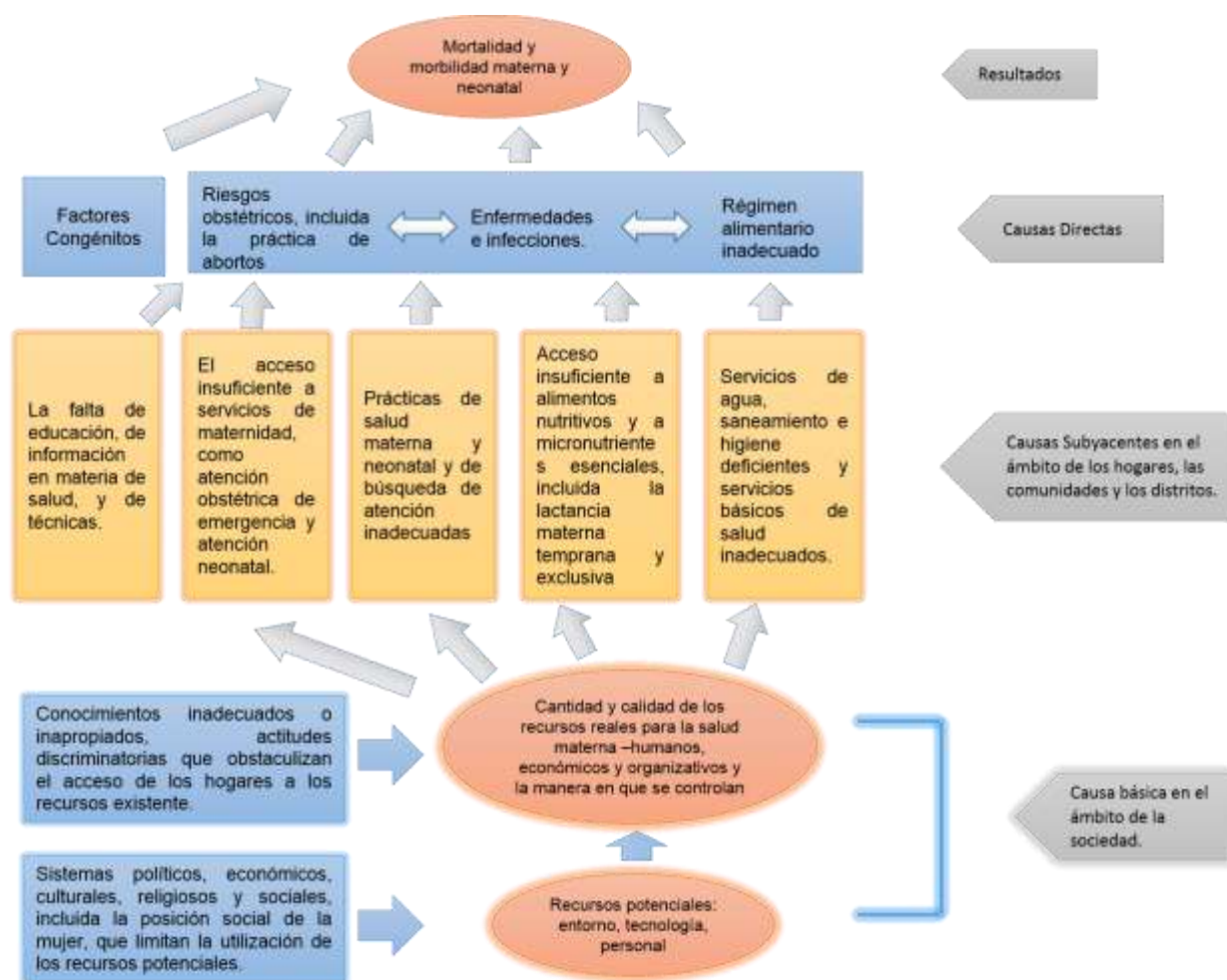


Figura 43. Proceso de morbilidad materna y neonatal

*Fuente Unicef.

Anexo 2. Cronograma de Actividades

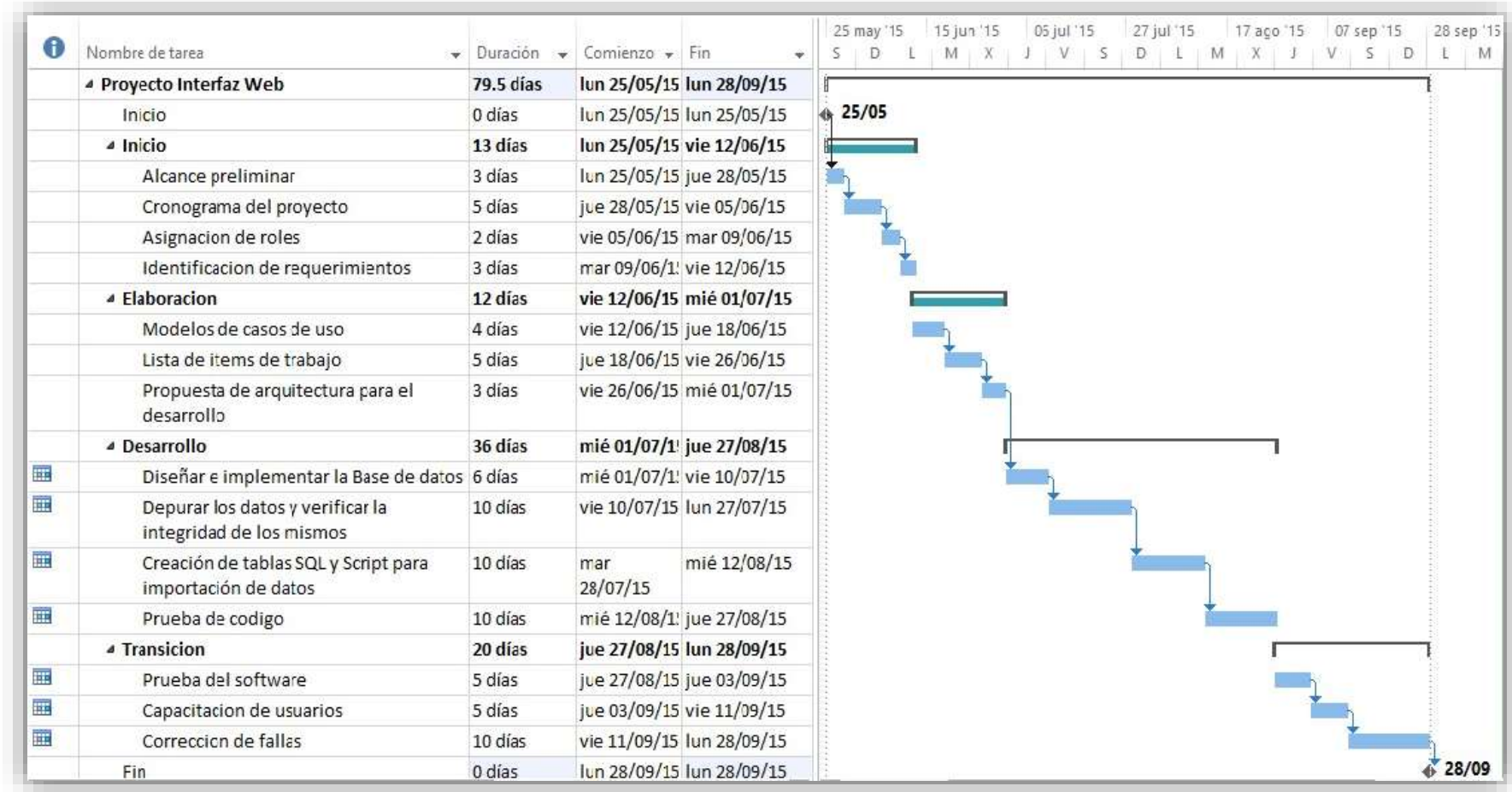


Figura 44 Cronograma de Proyecto

Anexo 3. Codificación Algoritmo de clasificación

```
namespace nBayes
{
using System;

publicclass Analyzer
{
privatefloat I = 0;
privatefloat invI = 0;

public Analyzer()
{
this.Tolerance = .05f;
}

publicfloat Tolerance { get; set; }

public CategorizationResult Categorize(Entry item, Index first, Index second)
{
float prediction = GetPrediction(item, first, second);

if (prediction <= .5f - this.Tolerance)
return CategorizationResult.Second;

if (prediction >= .5 + this.Tolerance)
return CategorizationResult.First;

return CategorizationResult.Undetermined;
}

publicfloat GetPrediction(Entry item, Index first, Index second)
{
foreach (string token in item)
{
int firstCount = first.GetTokenCount(token);
int secondCount = second.GetTokenCount(token);

float probability = CalcProbability(firstCount, first.EntryCount, secondCount,
second.EntryCount);

Console.WriteLine("{0}: [{1}] ({2}-{3}), ({4}-{5})",
token,
probability,
```

```

        firstCount,
        first.EntryCount,
        secondCount,
        second.EntryCount);
    }

float prediction = CombineProbability();
return prediction;
}

#region Private Methods

///<summary>
/// Calculates a probability value based on statistics from two categories
///</summary>
private float CalcProbability(float cat1count, float cat1total, float cat2count, float cat2total)
{
    float bw = cat1count / cat1total;
    float gw = cat2count / cat2total;
    float pw = ((bw) / ((bw) + (gw)));
    float
        s = 1f,
        x = .5f,
        n = cat1count + cat2count;
    float fw = ((s * x) + (n * pw)) / (s + n);

    LogProbability(fw);

return fw;
}

private void LogProbability(float prob)
{
    if (!float.IsNaN(prob))
    {
        I = I == 0 ? prob : I * prob;
        invI = invI == 0 ? (1 - prob) : invI * (1 - prob);
    }
}

private float CombineProbability()
{
return I / (I + invI);
}

```

```
#endregion
    }
}
```

Anexo 4. Algoritmo de obtención de Morbilidad

```
namespace nBayes
{
    using System;
    using System.Collections.Generic;
    using System.Linq;
    using System.Text.RegularExpressions;

    publicabstractclass Entry : IEnumerable<string>
    {
        public Entry()
        {
        }
    }

    publicstatic Entry FromString(string content)
    {
        returnnew StringEntry(content);
    }

    publicabstract IEnumerator<string> GetEnumerator();

    System.Collections.IEnumerator System.Collections.IEnumerable.GetEnumerator()
    {
        returnthis.GetEnumerator();
    }

    privateclass StringEntry : Entry
    {
        private IEnumerable<string> tokens;

        public StringEntry(string stringcontent)
        {
            tokens = Parse(stringcontent);
        }

        publicoverride IEnumerator<string> GetEnumerator()
        {
            return tokens.GetEnumerator();
        }
    }
}
```



```

///<summary>
/// Tokenizes a string
///</summary>
privatetatic IEnumerable<string> Parse(string source)
{
string clean = CleanInput(source);
string[] tokens = clean.Split(' ');
return tokens
    .Where(t => !t.Equals(" ", StringComparison.InvariantCultureIgnoreCase))
    .Select(t => t.ToLowerInvariant())
    .Distinct();
}

///<summary>
/// Replace invalid characters with spaces.
///</summary>
privatetaticstring CleanInput(string strIn)
{
return Regex.Replace(strIn, @"[\^\w\@-]", " ");
}
}
}
}
}

```

Anexo 5. Tabla comparativa de índice de masa corporal en el Perú.

Tabla 27. Comparativa de IMC en el Perú.

	Clasificación							
	Pesos(Kg.)							
	Delgadez			Normal	Sobrepeso	Obesidad		
IMC	III <16	II ≥16	I ≥17	≥18,5	≥25	I ≥30	II ≥35	III ≥40
1,35		29,1	30,9	33,7	45,5	54,6	63,7	72,9
1,36		29,5	31,4	34,2	46,2	55,4	64,7	73,9
1,37		30,0	31,9	34,7	46,9	56,3	65,6	75,0
1,38		30,4	32,3	35,2	47,6	57,1	66,6	76,1
1,39		31,9	32,8	35,7	48,3	57,9	67,6	77,2
1,40		31,3	33,3	36,2	49,0	58,8	68,6	78,4

1,41		31,8	33,7	36,7	49,7	59,6	69,5	79,5
1,42		32,2	34,2	37,3	50,4	60,4	70,5	80,6
1,43		32,7	34,7	37,8	51,1	61,3	71,5	81,7
1,44		33,1	35,2	38,3	51,8	62,2	72,5	82,9
1,45		33,6	35,7	38,8	52,5	63,0	73,5	84,1
1,46		34,1	36,2	39,4	53,2	63,9	74,6	85,2
1,47		34,5	36,7	39,9	54,0	64,8	75,6	86,4
1,48		35,0	37,2	40,5	54,7	65,7	76,6	87,6
1,49		35,5	37,7	41,0	55,5	66,6	77,7	88,8
1,50		36,0	38,2	41,6	56,2	67,5	78,7	90,0
1,51		36,4	38,7	42,1	57,0	68,4	79,8	91,2
1,52		36,9	39,2	42,7	57,7	69,3	80,8	92,4
1,53		37,4	39,7	43,3	58,5	70,2	81,9	93,6
1,54		37,9	40,3	43,8	59,2	71,1	83,0	94,8
1,55		38,4	40,8	44,4	60,0	72,0	84,0	96,1
1,56		38,9	41,3	45,0	60,8	73,0	85,1	97,3
1,57		39,4	41,9	45,6	61,6	73,9	86,2	98,5
1,58		39,9	42,4	46,1	62,4	74,8	87,3	99,8
1,59		40,4	42,9	46,7	63,2	75,8	88,4	101,1
1,60		40,9	43,5	47,3	64,0	76,8	89,6	102,4
1,61		41,4	44,0	47,9	64,8	77,7	90,7	103,6
1,62		41,9	44,6	48,5	65,6	78,7	91,8	104,9
1,63		42,5	45,1	49,1	66,4	79,7	92,9	106,2
1,64		43,0	45,7	49,7	67,2	80,6	94,1	107,5
1,65		43,5	46,2	50,3	68,0	81,6	95,2	108,9
1,66		44,0	46,8	50,9	68,8	85,6	96,4	110,2
1,67		44,6	47,4	51,5	69,7	83,6	97,6	111,5
1,68		45,1	47,9	52,2	70,5	84,6	98,7	112,8
1,69		45,6	48,5	52,8	71,4	85,6	99,9	114,2
1,70		46,2	49,1	53,4	72,2	86,7	101,1	115,6

1,71		46,7	49,7	54,0	73,1	87,7	102,3	116,9
1,72		47,3	50,2	54,7	73,9	88,7	103,5	118,3
1,73		47,8	50,8	55,3	74,8	89,7	104,7	119,7
1,74		48,4	51,4	56,0	75,6	90,8	105,9	121,1
1,75		49,0	52,0	56,6	76,5	91,8	107,1	122,5
1,76		49,5	52,6	57,3	77,4	92,9	108,4	123,9
1,77		50,1	53,2	57,9	78,3	93,9	109,6	125,3
1,78		50,6	53,8	58,6	79,2	95,0	110,8	126,7
1,79		51,2	54,4	59,2	80,1	96,1	112,1	128,1
1,80		51,8	55,0	59,9	81,0	97,2	113,4	129,6
1,81		52,4	55,6	60,6	81,9	98,2	114,6	131,0
1,82		52,9	56,3	61,2	82,8	99,3	115,9	132,4
1,83		53,5	56,9	61,9	83,7	100,4	117,2	133,9
1,84		54,1	57,5	62,6	84,6	101,5	118,4	135,4
1,85		54,7	58,1	63,3	85,5	102,6	119,7	136,9
1,86		55,3	58,8	64,0	86,4	103,7	121,0	138,3
1,87		55,9	59,4	64,6	87,4	104,9	122,3	139,8
1,88		56,5	60,0	65,3	88,3	106,0	123,7	141,3
1,89		57,1	60,7	66,0	89,3	107,1	125,0	142,8
1,90		57,7	61,3	66,7	90,2	108,3	126,3	144,4
1,91		58,3	62,0	67,4	91,2	109,4	127,6	145,9
1,92		58,9	62,6	68,1	92,1	110,5	129,0	147,4
1,93		59,5	63,3	68,9	93,1	111,7	130,3	148,9
1,94		60,2	63,9	69,6	94,0	112,9	131,7	150,5
1,95		60,8	64,6	70,3	95,0	114,0	133,0	152,1
1,96		61,4	65,3	71,0	96,0	115,2	134,4	153,6
1,97		62,0	65,9	71,7	97,6	116,4	135,8	155,2
1,98		62,7	66,6	72,5	98,0	117,6	137,2	153,8

Anexo 6. Datos obtenidos de una madre gestante menor de edad

Tabla 28. Datos de madre gestante menor de edad

Fecha	Comida	Cereales	Panes	Verduras y Hortalizas	Frutas	Aceites y Grasas	Productos Lácteos	Carnes Legumbres y Huevos	Complementos vitamínicos solido g	Agua	Descalcificantes	SAL	Azúcar	Condimentos
27/08/2015	Desayuno	1	3	1	1	0	0	1	0	2	0	1	5	1
27/08/2015	Media Mañana	0	0	0	1	0	0	1	1	2	0	0	15	0
27/08/2015	Almuerzo	0	0	2	1	2	0	1	0	1	0	2	3	1
27/08/2015	Media Tarde	0	0	0	1	0	0	0	0	2	0	0	15	0
27/08/2015	Cena	0	0	2	0	2	0	1	0	1	1	2	3	1
27/08/2015	Noche	1	3	1	1	0	0	1	0	2	0	1	5	1
28/08/2015	Desayuno	0	0	0	1	0	0	1	1	2	0	0	15	0
28/08/2015	Media Mañana	0	0	2	1	2	0	1	0	1	0	2	3	1
28/08/2015	Almuerzo	0	0	0	1	0	0	0	0	2	0	0	15	0
28/08/2015	Media Tarde	0	0	2	0	2	0	1	0	1	1	2	3	1
28/08/2015	Cena	1	3	1	1	0	0	1	0	2	0	1	5	1
28/08/2015	Noche	0	0	0	1	0	0	1	1	2	0	0	15	0
29/08/2015	Desayuno	0	0	2	1	2	0	1	0	1	0	2	3	1
29/08/2015	Media Mañana	0	0	0	1	0	0	0	0	2	0	0	15	0
29/08/2015	Almuerzo	0	0	2	0	2	0	1	0	1	1	2	3	1
29/08/2015	Media Tarde	1	3	1	1	0	0	1	0	2	0	1	5	1
29/08/2015	Cena	0	0	0	1	0	0	1	1	2	0	0	15	0
29/08/2015	Noche	0	0	2	1	2	0	1	0	1	0	2	3	1
30/08/2015	Desayuno	0	0	0	1	0	0	0	0	2	0	0	15	0
30/08/2015	Media Mañana	0	0	2	0	2	0	1	0	1	1	2	3	1

30/08/2015	Almuerzo	1	3	1	1	0	0	1	0	2	0	1	5	1
30/08/2015	Media Tarde	0	0	0	1	0	0	1	1	2	0	0	15	0
30/08/2015	Cena	0	0	2	1	2	0	1	0	1	0	2	3	1
30/08/2015	Noche	0	0	0	1	0	0	0	0	2	0	0	15	0
31/08/2015	Desayuno	0	0	2	0	2	0	1	0	1	1	2	3	1
31/08/2015	Media Mañana	1	3	1	1	0	0	1	0	2	0	1	5	1
31/08/2015	Almuerzo	0	0	0	1	0	0	1	1	2	0	0	15	0
31/08/2015	Media Tarde	0	0	2	1	2	0	1	0	1	0	2	3	1
31/08/2015	Cena	0	0	0	1	0	0	0	0	2	0	0	15	0
31/08/2015	Noche	0	0	2	0	2	0	1	0	1	1	2	3	1
01/09/2015	Desayuno	1	3	1	1	0	0	1	0	2	0	1	5	1
01/09/2015	Media Mañana	0	0	0	1	0	0	1	1	2	0	0	15	0
01/09/2015	Almuerzo	0	0	2	1	2	0	1	0	1	0	2	3	1
01/09/2015	Media Tarde	0	0	0	1	0	0	0	0	2	0	0	15	0
01/09/2015	Cena	0	0	2	0	2	0	1	0	1	1	2	3	1
01/09/2015	Noche	1	3	1	1	0	0	1	0	2	0	1	5	1
02/09/2015	Desayuno	0	0	0	1	0	0	1	1	2	0	0	15	0
02/09/2015	Media Mañana	0	0	2	1	2	0	1	0	1	0	2	3	1
02/09/2015	Almuerzo	0	0	0	1	0	0	0	0	2	0	0	15	0
02/09/2015	Media Tarde	0	0	2	0	2	0	1	0	1	1	2	3	1
02/09/2015	Cena	1	3	1	1	0	0	1	0	2	0	1	5	1
02/09/2015	Noche	0	0	0	1	0	0	1	1	2	0	0	15	0
03/09/2015	Desayuno	0	0	2	1	2	0	1	0	1	0	2	3	1
03/09/2015	Media Mañana	0	0	0	1	0	0	0	0	2	0	0	15	0
03/09/2015	Almuerzo	0	0	2	0	2	0	1	0	1	1	2	3	1

03/09/2015	Media Tarde	1	3	1	1	0	0	1	0	2	0	1	5	1
03/09/2015	Cena	0	0	0	1	0	0	1	1	2	0	0	15	0
03/09/2015	Noche	0	0	2	1	2	0	1	0	1	0	2	3	1
04/09/2015	Desayuno	0	0	0	1	0	0	0	0	2	0	0	15	0
04/09/2015	Media Mañana	0	0	2	0	2	0	1	0	1	1	2	3	1
04/09/2015	Almuerzo	1	3	1	1	0	0	1	0	2	0	1	5	1
04/09/2015	Media Tarde	0	0	0	1	0	0	1	1	2	0	0	15	0
04/09/2015	Cena	0	0	2	1	2	0	1	0	1	0	2	3	1
04/09/2015	Noche	0	0	0	1	0	0	0	0	2	0	0	15	0
05/09/2015	Desayuno	0	0	2	0	2	0	1	0	1	1	2	3	1
05/09/2015	Media Mañana	1	3	1	1	0	0	1	0	2	0	1	5	1
05/09/2015	Almuerzo	0	0	0	1	0	0	1	1	2	0	0	15	0
05/09/2015	Media Tarde	0	0	2	1	2	0	1	0	1	0	2	3	1
05/09/2015	Cena	0	0	0	1	0	0	0	0	2	0	0	15	0
05/09/2015	Noche	0	0	2	0	2	0	1	0	1	1	2	3	1