

**UNIVERSIDAD NACIONAL
“SANTIAGO ANTÚNEZ DE MAYOLO”**

**FACULTAD DE CIENCIAS
ESCUELA PROFESIONAL DE ESTADÍSTICA E
INFORMÁTICA**



**MINERÍA DE DATOS SOBRE LA OPINIÓN DE LA CALIDAD DEL
SERVICIO EDUCATIVO PERUANO Y LA LECTOESCRITURA
EN ESTUDIANTES DE EDUCACIÓN PRIMARIA DEL MÓDULO
EDUCACIÓN DE LA ENAHO TRIMESTRE 2, 2016.**

**TESIS PARA OPTAR EL TÍTULO DE:
LICENCIADO EN ESTADÍSTICA E INFORMÁTICA**

AUTORES:

Bach. MARCO ANTONIO ANAYA CARRILLO

Bach. JORGE LUIS BEDÓN TRUJILLO

ASESOR

Dr. ROGER PEDRO NORABUENA FIGUEROA

HUARAZ - ANCASH - PERÚ

2018

N° de Registro: T003



1. Datos del Autor: 1. Datos del Autor: 1. Datos del Autor:

Apellidos y Nombres: ANAYA CARRILLO MARCO ANTONIO

Código de alumno: 04.0163.7.UR

Teléfono: 943510533

Correo electrónico: marco.aycr@hotmail.com

DNI: 43952472

2. Modalidad de trabajo de investigación:

Trabajo de Investigación

Trabajo académico

Trabajo de suficiencia personal

Tesis

3. Título profesional o grado académico

Bachiller

Título

Segunda especialidad

Licenciado

Magister

Doctor

4. Título del trabajo de investigación

"MINERÍA DE DATOS SOBRE LA OPINIÓN DE LA CALIDAD DEL SERVICIO EDUCATIVO PERUANO Y LA LECTOESCRITURA EN ESTUDIANTES DE EDUCACIÓN PRIMARIA DEL MÓDULO EDUCACIÓN DE LA ENAHO TRIMESTRE 2, 2016."

5. Facultad de: Ciencias

6. Escuela, Carrera o programa: Escuela Profesional de Estadística e Informática.

7. Asesor:

Apellidos y Nombres: Dr. Norabuena Figueroa Roger Pedro

Teléfono: 968 850 116

Correo electrónico: nfrpedro_12@outlook.com

DNI: 41493243

A través de este medio autorizo a la Universidad Nacional Santiago Antúnez de Mayolo, publicar el trabajo de investigación en formato digital en el Repositorio Institucional Digital, Repositorio Nacional Digital de Acceso Libre (ALICIA) y el Registro Nacional de Trabajos de Investigación (RENATI).

Asimismo, por la presente dejo constancia que los documentos entregados a la UNASAM, versión impresión y digital, son las versiones finales del trabajo sustentado y aprobado por el jurado y son de autoría del suscrito en estricto respeto a la legislación en materia de propiedad intelectual

Firma: 

DNI:

FECHA:



1. Datos del Autor: 1. Datos del Autor: 1. Datos del Autor:

Apellidos y Nombres: BEDÓN TRUJILLO JORGE LUIS

Código de alumno: 04.0033.0.UR

Teléfono: 969252730

Correo electrónico: jorgebedon0406@gmail.com

DNI: 43592365

2. Modalidad de trabajo de investigación:

Trabajo de Investigación

Trabajo académico

Trabajo de suficiencia personal

Tesis

3. Título profesional o grado académico

Bachiller

Título

Segunda especialidad

Licenciado

Magister

Doctor

4. Título del trabajo de investigación

"MINERÍA DE DATOS SOBRE LA OPINIÓN DE LA CALIDAD DEL SERVICIO EDUCATIVO PERUANO Y LA LECTOESCRITURA EN ESTUDIANTES DE EDUCACIÓN PRIMARIA DEL MÓDULO EDUCACIÓN DE LA ENAHO TRIMESTRE 2, 2016."

5. Facultad de: Ciencias

6. Escuela, Carrera o programa: Escuela Profesional de Estadística e Informática.

7. Asesor:

Apellidos y Nombres: Dr. Norabuena Figueroa Roger Pedro

Teléfono: 968850116

Correo electrónico: nfrpedro_12@outlook.com

DNI: 41493243

A través de este medio autorizo a la Universidad Nacional Santiago Antúnez de Mayolo, publicar el trabajo de investigación en formato digital en el Repositorio Institucional Digital, Repositorio Nacional Digital de Acceso Libre (ALICIA) y el Registro Nacional de Trabajos de Investigación (RENATI).

Asimismo, por la presente dejo constancia que los documentos entregados a la UNASAM, versión impresión y digital, son las versiones finales del trabajo sustentado y aprobado por el jurado y son de autoría del suscrito en estricto respeto a la legislación en materia de propiedad intelectual

Firma: 

DNI:

FECHA:

DEDICATORIA

A la memoria de mi madre, María Susana (Q.E.P.D).

A mi padre Demetrio, mi hermana María y mi sobrina Melany por su cariño y amor.

A mi hijo Briam Anthony y mi esposa Lisbeth por regalarme todos los días, su amor sincero.

A toda mi familia.

Marco

A mis amados padres y hermanos quienes con sus palabras de aliento no me dejaban decaer para seguir adelante y siempre sea perseverante y cumpla con mis ideales.

Jorge

AGRADECIMIENTO

A nuestros padres y hermanos por su apoyo y sus consejos, a nuestros profesores de la facultad de ciencias, en especial de la escuela de Estadística e Informática por todas las horas de clase que son muy gratificantes.

Al Dr. Roger Pedro Norabuena Figueroa, nuestro asesor por su apoyo en la culminación de la presente investigación.

Marco y Jorge

HOJA DE VISTO BUENO

DR. ÁNGEL DECIDERIO MENDOZA LÓPEZ
PRESIDENTE

MSc. FERNANDO RAÚL ARCE ZÚÑIGA
SECRETARIO

DR. EDWIN JOHNY ASNATE SALAZAR
VOCAL

ÍNDICE

DEDICATORIA	ii
AGRADECIMIENTO	iii
HOJA DE VISTO BUENO	iv
ÍNDICE	v
ABSTRACT.....	viii
RESUMEN	ix
INTRODUCCIÓN	x
CAPÍTULO I: EL PROBLEMA.....	12
1.1. SELECCIÓN DEL PROBLEMA	13
1.2. JUSTIFICACIÓN DEL ESTUDIO	15
1.3. OBJETIVOS DE LA INVESTIGACIÓN	15
1.3.1. GENERAL.....	15
1.3.2. ESPECÍFICOS	16
CAPÍTULO II: MARCO TEÓRICO CONCEPTUAL	17
2.1. ANTECEDENTES DEL PROBLEMA.....	18
2.2. PLANTEAMIENTO TEÓRICO ATINGENTE.....	27

2.2.1. Bases teóricas	27
2.2.2. Definición de términos	71
CAPÍTULO III: HIPÓTESIS Y OPERACIONALIZACIÓN DE VARIABLES	72
3.1. Hipótesis	73
3.2. Operacionalización de variables	73
CAPÍTULO IV: METODOLOGÍA.....	74
4.1. Materiales y lugar de ejecución	75
4.1.1. Lugar	75
4.1.2. Materiales.....	75
4.1.3. Equipos.....	75
4.1.4. Recurso humano.....	75
4.2. Métodos	75
CAPÍTULO V: ANÁLISIS Y DISCUSIÓN DE RESULTADOS	78
5.1. Comprensión del problema.....	79
5.2. Comprensión de los datos	79
5.3. Preparación de los datos	81
5.4. Modelado	82
5.5. Evaluación	108
5.6. Implantación	109

5.7. Discusión de resultados	110
CAPÍTULO VI: CONCLUSIONES, LIMITACIONES Y RECOMENDACIONES ..	113
CONCLUSIONES	114
LIMITACIONES	115
RECOMENDACIONES.....	116
REFERENCIAS BIBLIOGRÁFICAS	117
ANEXOS	126

ABSTRACT

The present work of thesis according to its design is non-experimental descriptive type correlational and aimed to determine the relationship between the opinion of the quality Peruvian educational service and reading and writing in primary education students of the Education Module of the ENAHO quarter 2, 2016, through Mining of data. To do this, we worked with the database of the National Household Survey: education module for the second quarter of 2016.

The results obtained through data mining highlight the use of confirmatory factor analysis and decision trees. As observed in the confirmatory factor analysis, it is confirmed that the indicators of infrastructure, equipment, teaching of teachers, provision of educational materials and support for the participation of parents if they explain the quality of educational service in institutions of the country, while the results of the decision tree show that age influences the ability to read and write. Finally, the ability to read and write does not influence the qualification of the educational service received by students.

Keywords: data mining, reading and writing.

RESUMEN

El presente trabajo de tesis según su diseño es no experimental de tipo descriptivo correlacional y tuvo por objetivo determinar la relación entre la opinión de la calidad del servicio educativo peruano y la lectoescritura en estudiantes de educación primaria del Módulo Educación de la ENAHO trimestre 2, 2016, mediante la Minería de datos. Para ello se trabajó con la base de datos de la Encuesta Nacional de Hogares: módulo educación del segundo trimestre del 2016.

Los resultados obtenidos mediante la minería de datos destacan el uso del análisis factorial confirmatorio y los árboles de decisión. Según lo observado en el análisis factorial confirmatorio, queda confirmado que los indicadores de infraestructura, equipamiento, enseñanza de los maestros, dotación de los materiales educativos y el apoyo a la participación de los padres de familia si explican la calidad de servicio educativo en las instituciones del país, mientras que los resultados del árbol de decisión muestran que la edad influye en la capacidad de leer y escribir. Finalmente, la capacidad de leer y escribir no influye en la calificación del servicio educativo que reciben los estudiantes.

Palabras Clave: minería de datos, Leer y escribir.

INTRODUCCIÓN

El presente trabajo de tesis es el resultado final de mucha dedicación y esfuerzo en el filtrado, concatenación de bases de datos y procesamiento de los datos, para posteriormente ser analizados mediante la minería de datos, por intermedio del análisis factorial confirmatorio y los árboles de decisión.

Con el trabajo de investigación realizado se pretende conocer la relación entre la opinión de la calidad del servicio educativo peruano y la lectoescritura en estudiantes de educación primaria del Módulo Educación de la ENAHO trimestre 2, 2016, mediante la Minería de datos.

Para ello en el capítulo 1, se resalta la descripción del problema de estudio, se muestra en la justificación la importancia que tendrá este trabajo de investigación, y se plantean los objetivos que sirvieron de guías para la realización de la investigación.

En el capítulo 2, se muestran antecedentes, en la que resaltan los estudios de minería de datos y estudios referidos a la lectoescritura. Así mismo, se presenta de manera estructurada el desarrollo de la variable de estudio, con sus respectivas teorías que la fundamentan.

En el capítulo 3, se presenta la hipótesis del presente trabajo de investigación y se realiza la operacionalización de la variable estudiadas.

En el capítulo 4, se detallan aspectos metodológicos del presente trabajo de investigación. Así, el lugar de ejecución fue el INEI, con diseño de investigación de no experimental, de tipo descriptivo, con una población y muestra constituida por 29585 encuestas de la ENAHO del segundo trimestre del año 2016.

En el capítulo 5, se muestran los procedimientos paso a paso para la aplicación de la minería de datos, así como el análisis descriptivo respectivo de cada una de las características estudiadas.

Finalmente, en el capítulo 6, se muestra las conclusiones del presente trabajo de investigación y se sugieren recomendaciones que puedan ayudar a una mejor toma de decisiones.

CAPÍTULO I: EL PROBLEMA

1.1. SELECCIÓN DEL PROBLEMA

El Instituto nacional de Estadística e Informática (INEI), en el marco de sus actividades para la promoción y difusión de las investigaciones estadísticas que realiza, pone a disposición el sistema de Microdatos. Este sistema, proporciona las bases de datos y la documentación derivada de las investigaciones y encuestas ejecutadas por el INEI durante los últimos años, resguardando el secreto estadístico de la información.

Es así, que al contar con las bases de datos de la Encuesta Nacional de Hogares (ENAH) del segundo trimestre del año 2016, es necesario realizar minería de datos para reflejar la realidad peruana y difundirla.

Por otro lado, la calidad del servicio educativo peruano depende directamente del Ministerio de Educación, la cual se encuentra en permanente crisis en todos los niveles educativos de las instituciones educativas públicas dando como resultado la no existencia de un sistema de calidad educativa en los servicios educativos, dado que se carece de infraestructura acorde a la tecnología de hoy, deficiencias en el equipamiento de las aulas en cuanto a carpetas, laboratorios, computadoras, etc, deficiencias en cuando a la metodología de enseñanza de los docentes, carencia de materiales educativos como libros, láminas, maquetas, etc., y sobre todo la carencia de una escuela de padres donde se avoquen al apoyo y orientación de sus hijos.

Tal como lo afirmó en el diario El Comercio, en la entrevistas realizada a Hugo Díaz vicepresidente del Consejo Nacional de Educación titulada “los tres problemas que existen en el sector educación del Perú” considera como uno de dichos problemas es la falta de calidad y equidad, además indico que somos un país demasiado inequitativo en términos de calidad de los aprendizajes (Díaz, 2014).

Por otro lado, la curiosidad del niño por comprender el lenguaje escrito, genera la creación de hipótesis infantiles sobre cómo funciona este instrumento cultural. Al cuestionarse continuamente por esta actividad se hace consciente de los signos escritos, acercándose cada vez más a su dominio. La importancia de permitir que el niño se apropie de los conocimientos es evidente. Esta apropiación debe darse en la escuela y en el hogar como un proceso natural e implícito en todas las actividades cotidianas del niño. Promover la lectura expresando constantemente la funcionalidad de ésta en la cultura, aumenta la conciencia del manejo cotidiano del lenguaje escrito. Además, la escuela debe planear la instrucción con base en el nivel conceptual desarrollado espontáneamente por el niño y debe llevar a cabo actividades de enseñanza donde se promueva el auto-descubrimiento (Montealegre & Forero, 2006)

Por todo lo mencionado anteriormente, se plantea la siguiente pregunta de investigación: **¿Existe relación entre la opinión de la calidad del servicio educativo peruano y la lectoescritura en estudiantes de educación primaria del**

Módulo Educación de la ENAHO trimestre 2, 2016, mediante la Minería de datos?

1.2.JUSTIFICACIÓN DEL ESTUDIO

El presente trabajo de investigación es de gran importancia para el sector educación del país y en especial para la región Ancash, pues mediante la minería de datos se descubrió las posibles relaciones entre la opinión que tienen los estudiantes del nivel primaria respecto a la calidad del servicio educativo peruano que reciben en las instituciones educativas que estudian relacionado con su habilidad de leer y escribir correctamente.

Por otro lado, es de gran importancia para los investigadores del sector educación, pues tendrán a la mano información actual de la educación del nivel primaria en el Perú, así mismo, será de gran importancia para los profesionales de estadística, puesto que se hizo uso de la minería de datos para explorar la base de datos del módulo de educación de la ENAHO.

1.3.OBJETIVOS DE LA INVESTIGACIÓN

1.3.1. GENERAL

Determinar si existe relación entre la opinión de la calidad del servicio educativo peruano y la lectoescritura en estudiantes de educación primaria del Módulo Educación de la ENAHO trimestre 2, 2016, mediante la Minería de datos.

1.3.2. ESPECÍFICOS

- Realizar el análisis factorial confirmatorio sobre la opinión de la calidad del servicio educativo peruano en estudiantes de educación primaria del Módulo Educación de la ENAHO trimestre 2, 2016, mediante la Minería de datos.
- Analizar mediante los árboles de decisión la relación entre la opinión de la calidad del servicio educativo peruano y la lectoescritura en estudiantes de educación primaria del Módulo Educación de la ENAHO trimestre 2, 2016, mediante la Minería de datos.

CAPÍTULO II: MARCO TEÓRICO CONCEPTUAL

2.1. ANTECEDENTES DEL PROBLEMA

Montealegre & Forero (2006), en su trabajo de investigación titulado “Desarrollo de la lectoescritura: adquisición y dominio” realizado en Colombia, concluye que marca definitivamente el dominio del lenguaje escrito, pasar del momento de aprender a leer al momento en que se aprende leyendo. El estadio de dominio de la escritura se debe presentar al niño y a la niña como un paso en su camino hacia la inserción en la cultura; y se le debe también, resaltar su funcionalidad, su cotidianidad y su pertinencia para la comprensión de otros saberes. La escritura es un valioso objeto de estudio, pues precisamente la capacidad de producción escrita es la primera evidencia de la adquisición gradual de los principios del sistema alfabético, y más adelante una muestra sólida de un alto grado de apropiación de la lectoescritura. La posibilidad de transmitir conceptos e ideas a través de un texto escrito, implica la utilización de múltiples dominios con la claridad suficiente para transcribir de forma estructurada, clara y coherente las propias ideas. El contexto influye en el procesamiento del lenguaje escrito, pues la visión del mundo, la cultura, las interacciones sociales, la situación social y escolar, etc., determinan la estructura mental del sujeto, a la que integra la nueva información. Factores personales, sociales y ambientales son necesarios conocerlos antes de la estructuración de procedimientos metodológicos en la instrucción formal.

Correa Medina (2007), en su trabajo de tesis titulada “Conciencia fonológica y percepción visual en la lectura inicial de niños del primer grado de primaria” para optar el título de licenciado en psicología educacional de la Pontificia Universidad Católica del Perú,

concluye que el desempeño de los niños en conciencia fonológica se ubica en un estadio elemental, es decir, son capaces de segmentar palabras en sílabas e identificar rimas, pero aún les es difícil operar con fonemas. El desempeño de los niños en descodificación aún es lento, el 51.27% se ubica en un nivel por debajo de lo esperado. Sólo el 4.06% logra leer más de 50 palabras por minuto. El rendimiento de los niños en lectura aún no es inferencial, en su mayoría sólo llegan a alcanzar un nivel de comprensión literal. La conciencia fonológica y la percepción visual influyen positivamente en el rendimiento lector, tanto en rapidez como en comprensión. La conciencia fonológica posee una mayor relación con la rapidez y la comprensión de lectura que la percepción visual. No se apreciaron diferencias según género. En este sentido, tanto las niñas como los niños obtuvieron resultados similares en el puntaje global como en los subtests de las pruebas aplicadas.

Valerio Jiménez (2011), en su trabajo de investigación titulado “Problemas de lectoescritura” desarrollado en España, considera que los que trabajan en la docencia, saben lo duro que es a veces, que un determinado número de alumnos que, en muchas ocasiones, sin tener déficit mentales, sensoriales o incluso problemas de exclusión social o de estímulos ambientales, no rinda lo esperado o no aproveche todas las oportunidades que tiene y que seguramente arrastre durante todo su desarrollo educativo. Este alumnado puede presentar claras dificultades de lectoescritura, preocupando así a docentes y familiares. Los profesionales docentes, deben abordar estas dificultades, conociendo cuáles son las causas que inciden en la aparición de estas deficiencias, puesto que sin

“ayuda”, difícilmente podrán alcanzar los objetivos previstos para el curso, ciclo, etapa, etc., puesto que son primordiales las técnicas de lectura, escritura, comprensión, etc. Pero en muchas ocasiones no resulta fácil. La dislexia, dislalia, disgrafía y disortografía son algunas de las causas más frecuentes de dificultades en el aprendizaje lectoescritor y, por tanto, conviene conocer cómo actuar en los casos en los que se sospeche que puede existir alguna de estas alteraciones.

Perversi (2007), en su trabajo de tesis titulado “Aplicación de minería de datos para la exploración y detección de patrones delictivos en Argentina” para optar el grado de ingeniero industrial del Instituto Tecnológico de Buenos Aires Argentina, concluye que el presente trabajo ha demostrado no sólo que es factible aplicar minería de datos a la información criminal en Argentina, sino también su alto valor agregado para el análisis y la generación de nuevo conocimiento.

Conejo y Carmiol (2017), en su trabajo de investigación titulado “Conocimientos sobre la lectoescritura emergente y prácticas en las aulas para su promoción: Un estudio con docentes de Educación preescolar en Costa Rica” desarrollada en Costa Rica, resume que las concepciones y los conocimientos docentes afectan la práctica docente y esta, a su vez, influye en los logros de desarrollo en lectoescritura emergente de los niños preescolares. Los resultados indicaron diferencias según tipo de universidad en la cual se tituló la docente. Las docentes graduadas de universidades públicas otorgaron mayor importancia a prácticas efectivas en el aula para la promoción de la lectoescritura

emergente que las docentes graduadas de universidades privadas. Si bien la mayoría de participantes consideró la conciencia fonológica como un factor importante para el desarrollo de la lectoescritura, los conocimientos sobre esta fueron deficientes en toda la muestra. Igualmente, los conocimientos docentes sobre los fonemas y su habilidad para identificarlos en palabras concretas también fueron deficientes.

Cuetos, Molina, Suárez-Coalla y Llenderozas (2017), en su trabajo de investigación titulado “Validación del test para la detección temprana de las dificultades en el aprendizaje de la lectura y escritura”, desarrollado en España, concluyen que la prueba de detección temprana de las dificultades lectoras, que está siendo ampliamente utilizada, tanto por pediatras como por profesores de Educación Infantil, por su fácil y rápida aplicación, cuenta desde ahora con un sólido apoyo empírico. Aunque no es una prueba de diagnóstico de la dislexia, sí es capaz de alertar de posibles dificultades futuras en el aprendizaje de la lectura.

Sastre-Gómez, Celis-Leal, Roa-De La Torre y Luengas-Monroy (2017), en su trabajo de investigación titulado “La conciencia fonológica en contextos educativos y terapéuticos: efectos sobre el aprendizaje de la lectura”, desarrollado en Colombia, resumen que el aprendizaje de la lectura supone el desarrollo previo de una serie de habilidades, que van permitiendo a los niños la construcción y ganancia de destrezas en diferentes dimensiones hacia la consecución del progreso de sus procesos cognitivos. Entre ellas, se encuentra la conciencia fonológica, variable de mayor relevancia en cuanto al desarrollo lector se

refiere. La eficiencia de dicho mecanismo se relaciona directamente con el éxito en la alfabetización; las dificultades tanto para aprender a leer como para la lectura en sí misma, históricamente se han asociado a un déficit en el componente fonológico del lenguaje. Se evidencia que las distintas maneras en que, tanto el contexto pedagógico como el terapéutico, han generado estrategias facilitadoras que influyen en el normal desarrollo de las habilidades pre lectoras y de adquisición de la lectoescritura.

Magán-Hervás y Gértrudix-Barrio (2017), en su trabajo de investigación titulado “Influencia de las actividades audio-musicales en la adquisición de la lectoescritura en niños y niñas de cinco años”, desarrollado en España, afirman que la música sí influye favorablemente en el desarrollo de la lectoescritura. Las prácticas auditivas sonoras musicales, relacionadas con el reconocimiento de sonidos (memoria auditiva), son más eficaces para el desarrollo de la lectoescritura. Además, las canciones infantiles también favorecen este proceso de aprendizaje, por su grado implícito en la motivación del niño y de la niña.

Palos, Ávalos, Flores y Montes (2017), en su trabajo de investigación titulado “Creencias de madres y docentes sobre el aprendizaje de la lectoescritura en Educación Preescolar” desarrollado en Costa Rica, resumen que el aprendizaje de la lectoescritura es un logro primordial que plantea todo sistema educativo, constituye un prerrequisito importante en el desarrollo de habilidades cognitivas complejas que permiten a las personas hacer frente a exigencias de la vida cotidiana. El aprendizaje de la lectoescritura son las interacciones

en la casa y en la escuela y desempeñan un papel decisivo en el desarrollo de habilidades para la lectoescritura y la competencia lectora. Los hallazgos permiten afirmar que la visión del jardín de niños como espacio de entretenimiento y cuidado ha sido superada; sin embargo, respecto a los procesos de lectoescritura existe confusión y una tendencia a reducirlos simplemente a la decodificación. Finalmente, la importancia de identificar y desarrollar modelos que favorezcan estos procesos y la necesidad de que las madres de familia colaboren corresponsablemente para asegurar el acceso y participación exitosa en iniciativas que impulsen el desarrollo de habilidades para la lectoescritura.

Hincapié (2017), en su trabajo de investigación titulado “Didáctica de la lectoescritura universitaria: dos reseñas a Jonathan Potter” desarrollado en Colombia, concluye que la escritura y la lectura constituyen dos procesos estrechamente vinculados e invaluable en el quehacer de la vida cotidiana de los estudiantes y profesores universitarios, ya que posibilitan la construcción del pensamiento y la formación de los individuos en este ámbito académico. De igual forma, debido a las características de la lectura y de la lectura descriptiva, puede afirmarse que la representación que opera en la mayoría de alumnos en el momento de ingreso [a la universidad] no contribuye al desarrollo de competencias de lectura y escritura académicas, y que, más allá de las distancias observadas, esas representaciones guían las acciones de los alumnos de modo que los llevan a repetir experiencias conocidas en otro ámbito, tratando de adecuarlas al nuevo, pero sin éxito.

Soto (2017), en su trabajo de investigación titulado “El cuento como mediación pedagógica para el fortalecimiento de la lectoescritura” desarrollado en Colombia, concluye que el uso de cuentos como herramienta pedagógica es conveniente para el fortalecimiento de la comprensión lectora, en la medida en la que se hacen parte del medio diseñado por el educador para dinamizar el proceso de enseñanza-aprendizaje. En efecto, se consideró lo propuesto por Montessori, al ratificarse que: el educador diseña el medio con el fin de satisfacer las necesidades de autoconstrucción del niño, pero debe abstenerse de ser partícipe de este proceso y más bien, asumirse como una “sombra”, de manera que los niños aprendan por sí solos. En este sentido, se puso de relieve el enfoque de aprendizaje constructivista, en la medida en que el conocimiento no sólo hacía parte de un proceso de transferencia lineal docente-estudiante, sino que se conformaba un proceso de construcción temática, en el que el educador era un guía y el estudiante, el dueño de su propio aprendizaje. En síntesis, el fortalecimiento de las competencias lectoras a través de una metodología ideo-visual de aprendizaje usando cuentos como herramienta pedagógica constituye una forma distinta de hacer pedagogía. Sin duda, la utilización de libros, revistas y guías escolares, a pesar de estar referenciadas en los planes curriculares de los docentes, constituyen formas monótonas de enseñar lectura y propician el enfoque conductista y memorístico del proceso de enseñanza-aprendizaje. De esta manera, hacer uso de los cuentos y las dramatizaciones de historias, además de los conversatorios de aula y los juegos de roles, como parte del apoyo a las acciones pedagógicas, descongela, por así decirlo, la evolución de la pedagogía y la didáctica del aprendizaje, posibilitando la entrada en vigencia de nuevos recursos para enseñar y aprender. Definitivamente,

permitirle al estudiante observar, jugar, interactuar, escuchar y vivir historias, involucra, sin duda, nuevas experiencias para el educador, dotándole de herramientas ideo-visuales y de nuevos conocimientos sobre la enseñanza de lectura en el aula.

Mendoza y Rojas (2016), en su trabajo de investigación titulado “Resultados concluyentes del estudio sobre el nivel de lectoescritura en cadetes de la Escuela Militar (2010-2014)” desarrollado en Bogotá - Colombia, concluyen que contar con un buen manejo de la ortografía es fundamental en el desarrollo profesional y personal del cadete de la Escuela Militar, ya que gran parte del quehacer diario pasa por relaciones a través del medio escrito, por lo tanto, es de suma importancia aprender a organizar ideas y expresarlas de modo claro. Redactar correctamente escritos también implica impartir instrucciones claras, concisas, directas y, así, evitar confusiones y malos entendidos. Una buena comunicación escrita, respetando el uso adecuado de las reglas ortográficas, provocará en el lector la correcta interpretación del texto, despertará interés y propiciará un mayor grado de entendimiento, que se traducirá en una respuesta eficaz. Los programas de talleres creativos, generan cambios en el programa de lectoescritura incrementando los niveles inferencial y crítico. Así, los profesores deben incentivar más la lectura y la forma correcta de escribir las palabras, aprovechando el plan lector.

Castillo, Medina y Vega (2016) en su trabajo de investigación titulado “Esquema de Visualización para Modelos de Clústeres en Minería de Datos” desarrollado en Chile,

concluyen que las combinaciones de técnicas de minería de datos mejoran la representación visual del clúster siendo más entendibles por los usuarios.

Estrada, Zamarripa, Zúñiga y Martínez (2016), en su trabajo de investigación titulado “Aportaciones desde la minería de datos al proceso de captación de matrícula en instituciones de educación superior particulares” desarrollado en Costa Rica, concluyen que el desarrollo de nuevas tecnologías de información para mejorar el proceso de captación de matrícula es indispensable, esto debido a la reducción del mercado que obliga a las instituciones educativas a dirigir su mirada hacia técnicas de minería de datos para apoyar este proceso, ya que la minería de datos es una herramienta fundamental que debe usarse en los ambientes universitarios con la finalidad de hacer más eficiente sus procesos, aunque es indispensable recordar que no sustituye en ningún momento la experiencia y el contacto humano que es el centro de los procesos de promoción y captación, ya que tan sólo brindan una base más objetiva para la toma de decisiones.

Visbal, Mendoza y Orjuela (2017), en su trabajo de investigación titulado “Predicción de la eficiencia de las instituciones de educación superior colombianas con análisis envolvente de datos y minería de datos” desarrollado en Colombia, concluyen que el uso del análisis envolvente de datos conjuntamente con herramientas de minería de datos indica que se logra una mejora sustancial en el análisis de eficiencia y clasificación de los datos.

Ochoa, Orellana, Sánchez y Dávila (2014), en su trabajo de investigación titulado “Componente web para el análisis de información clínica usando la técnica de Minería de Datos por agrupamiento” desarrollado en Cuba, concluyen que la Minería de Datos es un proceso eficaz para dar respuestas a preguntas complejas de Inteligencia de Negocios, y es por ende, una buena manera de convertir datos en información, y esta a su vez en conocimiento, para la correcta toma de decisiones

2.2. PLANTEAMIENTO TEÓRICO ATINGENTE

2.2.1. Bases teóricas

Lectoescritura

La lectoescritura constituye un pilar básico en la vida de todo ser humano. No solo se trata de una herramienta indispensable para acceder a los objetivos y contenidos educativos si no que es uno de los principales vehículos de la cultura existente (Lucas, 2014). La importancia de la lectura y escritura radica en que se trata de un agente imprescindible en la trasmisión cultural entre generaciones y personas en general, siendo el vehículo que permite adquirir información, cultura y aprendizajes de manera universal.

La lectura y la escritura no sólo permiten desarrollar la competencia lectora, si no que a través de las mismas se establece una estrecha relación con el resto de competencias básicas, especialmente con la competencia para aprender a aprender y el tratamiento

de la información y competencia digital, permitiendo así la adquisición de la alfabetización mediática. (BOCYL, 2011).

La lectoescritura constituye una estrategia y un proceso en nuestros alumnos; como estrategia en el proceso de enseñanza-aprendizaje, y como proceso para acercarnos a la comprensión del texto. “La lectura y la escritura son elementos inseparables de un mismo proceso mental” (Ferreiro, 2011)

La Real Academia de la Lengua Española (RAE) define la lectura como la acción de leer, la cual consiste en pasar la vista por un escrito o impreso comprendiendo la significación de los caracteres empleados.

Dubois (2000) define: “la lectura es el producto de la interacción entre el pensamiento y el lenguaje”. Así pues, si el lector extrae el significado del texto y el propósito del autor, entonces comprenderá dicho texto.

Alvarez (1993) distingue tres tipos de lectura: oral, en voz alta y visual. En la lectura oral el principal destinatario es nuestro propio cerebro. Su finalidad es que el lector las escuche para su decodificación

La Real Academia de la Lengua Española (RAE) define escribir como la representación de palabras o ideas con letras u otros signos trazados en papel u otra superficie.

El profesor Robert Rigal (2006) define la escritura como la actividad perceptivo-motriz en la que la coordinación visomanual y la intervención de los músculos y articulaciones, los cuales se encargan de la motricidad fina de la mano, juegan un papel primordial. Así, toda grafía necesita el desplazamiento de la mano para formar las letras y palabras y una percepción visual que requiere la intervención de diferentes grupos musculares, de ahí que se defina como una actividad perceptivomotriz.

Silva (1997) considera que la lectura se constituye en la principal dieta nutricional para el crecimiento creativo del individuo, principalmente cuando relacionamos la lectura al contexto social.

Orlandi (2001) entiende la lectura en un sentido amplio como la necesidad de comprender su característica polisémica en lo que concierne a su índole conceptual. Por lo tanto, es importante distinguir los varios sentidos con los que se entiende a la lectura, por ello sea empleada de manera indiferente tanto en la escritura como en la oralidad.

Martins (2004) considera que conviene asimilar que las innumerables conceptualizaciones vigentes sobre la lectura pueden ser sintetizadas en dos caracterizaciones:

- Como decodificación mecánica de signos lingüísticos, mediante el aprendizaje establecido partiendo del condicionamiento estímulo respuesta.
- Como proceso de entendimiento copioso, cuya dinámica involucra componentes sensoriales, emocionales, intelectuales, fisiológicos, neurológicos, así como culturales, económicos y políticos.

Solé (1998) afirma que: “la lectura es el proceso mediante el que se comprende el lenguaje escrito”. En dicha comprensión intervienen tanto el texto, su forma y el contenido, como el lector, sus expectativas y conocimientos previos.

Italo y Vargas (1995), afirma que: Tengo la seguridad de que la lectura no es comparable a cualquier otro medio de aprendizaje y de comunicación, porque ella posee un ritmo que es gobernado por la avidez del lector; la lectura abre un campo de interrogantes, de meditación y de análisis crítico, ósea, de libertad; la lectura es una correspondencia no apenas con el libro, pero también con nuestro mundo interior que el libro desvela.

En la lectura la comprensión está estrechamente relacionada a la necesidad del autor y del lector en hablar la misma lengua, con el objetivo de llegar juntos al vasto campo del entendimiento; ya que este es el resultado de una compleja actividad en la que el lector debe ejercitar las operaciones y las habilidades relativas a la comprensión lectora y recurrir a muchos tipos de conocimientos, como aclara Colomer (2002).

En el diálogo educativo existen innumerables divergencias teóricas, pero ninguna niega la importancia de aprender a leer y ni la de todas las implicaciones para garantizar el desarrollo cultural del alumno. Por ello, es uno de los principales roles imputados a la escuela el de formar a buenos lectores y caso esto no acontezca la institución deja de adjudicarse su papel representativo de penetración en los horizontes de la cultura, una vez que “la mayor parte del conocimiento que alguien puede obtener en su vida será mediante la lectura, aunque sea fuera de la escuela.” Ferreiro en Zorzi (2003).

Abordar a la lectura bajo la perspectiva de desarrollar el comportamiento lector, implica en traer a la luz de la discusión el contexto experimentado durante el desarrollo de las prácticas pedagógicas de la lectura, que desafía a los profesores a transponer conceptos arraigados sobre una enseñanza tradicional, que profesa a la centralización docente como modo de promocionar y controlar el avance de la comprensión del sistema de la lectura por parte del niño, para así entenderla como sujeto conocedor, o

sea, que busca activamente el conocimiento del mundo a su alrededor. Según Carvajal y Ramos (2001). Así, Para algunos profesores la finalidad de la enseñanza de aprender a leer, es la de que los alumnos aprendan a leer, a explicar, a prevenir, a responder preguntas realizadas después de la lectura o a verificar los hechos de la realidad, en este caso la enseñanza de la lectura ocurre de manera fragmentada y ordenada. Al inicio del proceso de enseñanza es concebido como una habilidad y no como un conocimiento, y son programadas actividades de tipo visual, motora o auditiva, en forma de copia, imitación y preparación.

La escritura entendida como señales, se define como la acción motriz sobre cualquier material, ya sea dibujo, pintura, grabado o rayado. El uso de herramientas deja ver claramente el adelanto técnico que implicaba la escritura. Este tipo de señales son imágenes plásticas visibles que pueden ser divididas en: pictórico y escritura. El tipo pictórico es aquel que reproduce el mundo independientemente del lenguaje, a diferencia de la escritura que con formas pictóricas que se convierten en símbolos, dan paso a nociones de valor lingüístico. Según Ignace J. Gelb (1987) es posible hablar de símbolos con valor lingüísticos ya que estos al manifestarse omiten detalles artísticos y expresan lo más exacto posible lo que se quiere transmitir.

Tal y como contempla Braslavsky (1991) en su obra “Enseñar a entender lo que se lee”, las investigaciones realizadas sobre la alfabetización temprana, tanto en familias

con miembros letrados como en familias con miembros no letrados, confirman la importancia de tienen los padres, los hermanos mayores y los maestros en el aprendizaje de la lectura y escritura. Dicho aprendizaje no sólo se realiza en el centro escolar, sino en muchos ambientes y contextos externos a la escuela. Las actividades que se realizan en familia relacionadas con la lectura y escritura en edades tempranas asientan las bases para que los niños aprendan a leer y escribir en la escuela, así como el placer por la lectura que los padres comparten con sus hijos fomenta actitudes positivas hacia la lectura. Posteriormente, será el tiempo de ocio y el placer por la lectura lo que hará que se perfeccionen y aumenten las habilidades lectoras. El ambiente donde se desenvuelve el niño debe estar relacionado con la lectura, es decir, en su entorno cuenten con la presencia de libros, revistas, documentos... y el hecho que la lectura sea una actividad placentera y cotidiana en su vida.

Calidad de los servicios educativos

La calidad del diseño para Municio (2000) es una medida de ajuste del producto o servicio a los requisitos demandados por los clientes. No es de calidad una silla por sus materiales, sino por su adecuación para ver la televisión o para trabajar en la oficina. La diferencia está en el diseño. Un programa de matemáticas (diseño de una ejecución futura) puede ser válido (en cuanto a calidad) para alumnos de estilo cognitivo convergente pero inadecuado para alumnos divergentes.

Según Edwards Deming (2000) calidad es todo aquello que le hiciera ahorrar a la empresa y a su vez cumplir a tiempo con la entrega del producto al cliente, donde esto cambió totalmente la manera de pensar a todos los gerentes al decirles que era más barato hacer un producto nuevo que corregirlo ya que esto conlleva a la pérdida de tiempo y esfuerzo, también Deming recalcó en observar las líneas de producción para observar donde se estaba dando el problema y por qué y así eliminar lo que estaba fallando de raíz y esto nos ayudaba a conocer cada vez mejor el sistema y aplicarlo a una forma de mejora continua y contribuir a la satisfacción del cliente.

Para Juran (1967), la calidad es "La adecuación para el uso satisfaciendo las necesidades del cliente".

Kaoru Ishikawa (1988) define a la calidad como: "Desarrollar, diseñar, manufacturar y mantener un producto de calidad que sea el más económico, el útil y siempre satisfactorio para el consumidor".

Se busca la calidad en la conformidad entre los requisitos o especificaciones del diseño y el producto o servicio acabado. Según Villarroel (1997), lograr esto supone la existencia de requisitos de diseño y la condición de ser sobrepasados en la realización. Pero, además implica haber realizado bien las tareas adecuadas y no haber realizado

tareas inadecuadas. Ésta es la única etapa de calidad específicamente interna y su conexión con el cliente viene a través del diseño y, posteriormente por su aceptación por el destinatario.

Lo más importante es saber diferenciar lo que es calidad de lo que no lo es. Crosby (1980), señala que la calidad de un bien o servicio no es otra cosa que la conformidad del mismo con los requerimientos de los clientes. Este concepto hace énfasis en el usuario del bien o servicio prestado: el nivel de calidad corresponde a una percepción individual mientras el concepto es genérico. Juran la refiere como la adecuación de uso de un bien o servicio, es decir, hasta qué punto ese bien o servicio satisface las necesidades del usuario. Por su parte, Feigenbaum y Harrington, citados por Crosby (1980), la conceptúan en términos del nivel de cumplimiento o superación de las expectativas de los clientes o usuarios.

Berry (2002), señala: "...la calidad es la satisfacción de las necesidades de los clientes" (p. 2). Esta definición está centrada en los resultados del proceso de logro de la calidad, en los efectos de su aplicación y en los beneficiarios de los servicios y/o productos: el cliente. Un concepto ampliado establece que la calidad consiste en satisfacer las necesidades de los clientes y sus expectativas razonables.

La educación tiene como propósito la formación del hombre, tanto en su aspecto técnico como humanístico. Quiere decir esto que lo que se persigue es no solo formar un individuo capaz de enfrentar, con inteligencia los problemas de la vida real y darles soluciones efectivas y concretas utilizando o desarrollando la tecnología más adecuada a las necesidades, sino también que ese individuo esté preparado para entender a sus semejantes y vivir en sociedad, con una conducta basada en los más nobles principios morales. En ese mismo sentido la Comunidad Europea (1995, p.4), ha señalado “...la educación y la formación tienen por función esencial la integración social y el desarrollo personal, mediante la asunción de valores comunes; la transmisión de un patrimonio cultural y el aprendizaje de la autonomía”. Esto implica una formación integral del hombre.

Para alcanzar estas metas es necesaria y conveniente una educación de calidad. Esto significa muchas cosas para diferentes personas. Cada uno tiene un concepto propio de lo que eso es y una visión personal de cómo lograrla. En este sentido, Cantón (2000: 61), afirma que “...la educación siempre será lo que sea la sociedad en la que se desenvuelva”. Por su parte, García Hoz (2000: 3) sostiene que la calidad de la educación viene determinada por dos componentes, de una parte, por la disposición, en la que se incluyen las condiciones de integridad y coherencia y, por la otra, la eficacia. En tal sentido define la calidad educativa como “...el modo de ser de la educación que reúne las condiciones de integridad, coherencia y eficacia. En otras

palabras, una educación tiene calidad en la medida en que es completa, coherente y eficaz”.

La eficacia, de acuerdo a García Hoz (2000), se refiere a la necesidad de que todos los componentes cumplan adecuadamente su función, de tal forma que el ser humano desarrolle todas sus potencialidades y compense sus limitaciones. La eficacia no refiere, en este caso, a la producción abundante de algo material, refiere más bien a la indicación de que una actividad ha alcanzado el fin que persigue. La eficacia opera sobre la aceptación previa de una finalidad que se deben concretar en objetivos que han de ser alcanzados utilizando los medios que se puedan disponer. La eficacia es, entonces, una relación entre objetivos (extensión, nivel, dificultad) que han de ser alcanzados; los medios de que se dispone (recursos materiales, humanos, tecnológicos) y la organización y métodos utilizados en la actividad misma de educar y los resultados, que son las metas finales a alcanzar.

Una mejor calidad educativa asegura un mayor ingreso de por vida para aquellas personas que la reciben y asegura un crecimiento rápido y sostenible para aquellos países que la proveen, ya que una mejora en la calidad de la educación aumenta las tasas de innovación e invención y asegura la productividad mediante mejores métodos de producción y tecnología.

Entre los principales factores que determinan la calidad de la educación podemos mencionar los siguientes:

- Los recursos materiales disponibles: aulas de clase, aulas de recursos, biblioteca, laboratorios, patio, instalaciones deportivas, mobiliario y recursos educativos.
- Los recursos humanos: nivel científico y didáctico del profesorado, experiencia y actitudes del personal en general, capacidad de trabajar en equipo, ratios alumnos/profesor, tiempo de dedicación. Los servicios y las actuaciones que realizan las personas son los que determinan la calidad de toda organización. En este sentido es muy importante su participación y compromiso.
- La dirección, gestión administrativa y académica del centro: labor directiva, organización, funcionamiento de los servicios, relaciones humanas, coordinación y control.
- Aspectos pedagógicos: PEI (Proyecto Educativo Institucional), POA (Proyecto Operativo Anual), evaluación inicial de los alumnos, adecuación de los objetivos y los contenidos, tratamiento de la diversidad, metodología didáctica, utilización de los recursos educativos, evaluación, tutorías, logro de los objetivos previstos.

Según el Consejo Nacional de Educación (2007), la mejora de los aprendizajes en la educación básica depende de un conjunto de factores, que tienen que ver con el equipamiento de las instituciones, la gestión escolar, la formación de los profesores, nuevas reglas de juego para el ejercicio de la docencia, evaluación periódica de los progresos en el rendimiento de los alumnos, una educación inicial con mayor cobertura y calidad y programas que compensen las desventajas iniciales de los alumnos a nivel de su salud y nutrición. Todos estos aspectos son prioritarios y están explícitamente demandados en el Proyecto Educativo Nacional. Pero en todos los casos, un mecanismo capaz de asegurar al docente y al director apoyo y asistencia pedagógica permanente a sus esfuerzos de cambio, resulta indispensable.

En ese sentido, cuando existe un buen servicio, es decir, en un entorno de calidad educativo, los niños logran desarrollar sus potencialidades. Al respecto, Martí (2003) señala que hay un amplio consenso en considerar que los niños que se desarrollan en entornos alfabetizados incursionan en el conocimiento de las características formales, aunque con distintos ritmos y con relevantes variaciones según el grado y modos en que se emplean en los entornos cotidianos y educativos de los niños.

Los niños logran aprendizajes de acuerdo a los servicios que brindan las instituciones educativas, en un contexto social y cultural. Esto significa que la institución educativa, al brindar servicio de calidad significa que se le brinda las oportunidades para que los

niños logren aprendizajes significativos acorde a actividades también significativas. Al respecto, Vosniadou (2006) señala que muchas actividades escolares no son significativas dado que los estudiantes no entienden por qué las están haciendo ni cuáles son su propósito y utilidad. Algunas veces las actividades escolares no son significativas porque culturalmente son inapropiadas. Muchas escuelas son comunidades en las que niños de diversas culturas aprenden juntos. Hay diferencias culturales sistemáticas en las prácticas, hábitos, roles sociales, etcétera, que influyen en el aprendizaje. Algunas veces, actividades significativas para estudiantes que provienen de un grupo cultural no lo son para alumnos que provienen de otro grupo cultural.

Minería de datos

La minería de datos o exploración de datos (es la etapa de análisis de "Knowledge Discovery in Databases" o KDD) es un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos.

Se denomina Minería de Datos [Servente & García-Martínez, 2002; Perichinsky & García-Martínez, 2000; Perichinsky et al., 2000; Perichinsky et al., 2001; Perichinsky et al., 2003] al conjunto de técnicas y herramientas aplicadas al proceso no trivial de

extraer y presentar conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con objeto de predecir de forma automatizada tendencias y comportamientos; y describir de forma automatizada modelos previamente desconocidos [Piatetski-Shapiro et al., 1991; Chen et al., 1996; Mannila, 1997]. El término Minería de Datos Inteligente [Evangelos & Han, 1996; Michalski et al., 1998] refiere específicamente a la aplicación de métodos de aprendizaje automático [Michalski et al., 1983; Holsheimer & Siebes, 1991], para descubrir y enumerar patrones presentes en los datos, para estos, se desarrollaron un gran número de métodos de análisis de datos basados en la estadística [Michalski et al., 1982]. En la medida en que se incrementaba la cantidad de información almacenada en las bases de datos, estos métodos empezaron a enfrentar problemas de eficiencia y escalabilidad y es aquí donde aparece el concepto de minería de datos. Una de las diferencias entre el análisis de datos tradicional y la minería de datos es que el primero supone que las hipótesis ya están construidas y validadas contra los datos, mientras que el segundo supone que los patrones e hipótesis son automáticamente extraídos de los datos [Hernández Orallo, 2000].

La Minería de Datos es un proceso completo de descubrimiento de conocimiento que involucra varios pasos [Morales, 2003]:

- Entendimiento del dominio de aplicación, el conocimiento relevante a utilizar y las metas del usuario.

- Seleccionar un conjunto de datos en donde realizar el proceso de descubrimiento.
- Limpieza y pre-procesamiento de los datos, diseñando una estrategia adecuada para manejar ruido, valores incompletos, valores fuera de rango, valores inconsistentes, etc..
- Selección de la tarea de descubrimiento a realizar, por ejemplo, clasificación, agrupamiento o clustering, reglas de asociación, etc..
- Selección de los algoritmos a utilizar.
- Transformación de los datos al formato requerido por el algoritmo específico de explotación de datos, hallando los atributos útiles, reduciendo las dimensiones de los datos, etc..
- Llevar a cabo el proceso de minería de datos para encontrar patrones interesantes.
- Evaluación de los patrones descubiertos y presentación de los mismos mediante técnicas de visualización. Quizás sea necesario eliminar patrones redundantes o no interesantes, o se necesite repetir algún paso anterior con otros datos, con otros algoritmos, con otras metas o con otras estrategias.
- Utilización del conocimiento descubierto, ya sea incorporándolo dentro de un sistema o simplemente para almacenarlo y reportarlo a las personas interesadas.

Es muy importante la etapa del pre-procesamiento de los datos y su transformación al formato requerido por el algoritmo, ya que dependiendo de cómo se realicen estas

tareas, va a depender la calidad final de los patrones descubiertos. Un patrón es interesante si es fácilmente entendible por las personas, potencialmente útil, novedoso o valida alguna hipótesis que el usuario busca confirmar. Un patrón interesante representa conocimiento [Ale, 2005].

Las principales técnicas de minería de datos se suelen clasificar según su tarea de descubrimiento en: Agrupación o clustering, clasificación y asociación.

Agrupación de datos

La agrupación o clustering consiste en agrupar un conjunto de datos basándose en la similitud de los valores de sus atributos. El clustering identifica regiones densamente pobladas, denominadas clusters, de acuerdo a alguna medida de distancia establecida [Chen et al., 1996]. De esta manera se busca maximizar la similitud de las instancias en cada cluster y minimizar la similitud entre clusters [Han & Kamber, 2001].

La técnica de clustering ha sido estudiada en las áreas de la estadística [Cheeseman & Stutz, 1996; Jain & Dubes, 1988], machine learning [Fisher, 1996], base de datos espaciales y minería de datos [Cheeseman & Stutz, 1996; Ester et al., 1995; Ng & Han, 1994; Zhang et al., 1996].

Dos de los algoritmos de clustering más utilizados son Self Organizing Maps (SOM) y K-means. SOM, también denominado redes de Kohonen, fue creado por Teuvo Kohonen en 1982. Se trata de un modelo de red neuronal con capacidad para formar mapas de características de manera similar a como ocurre en el cerebro. SOM está basado en el aprendizaje no supervisado y competitivo, lo cual quiere decir que no se necesita intervención humana durante el mismo y que se necesita saber muy poco sobre las características de la información de entrada. SOM provee un mapa topológico de datos, que se representan en varias dimensiones, utilizando unidades de mapa (las neuronas) para simplificar la representación [Kohonen, 1995]. Las neuronas usualmente forman un mapa bidimensional, por lo que el mapeo transforma un problema de muchas dimensiones en el espacio, a un plano. La propiedad de preservar la topología significa que el mapeo preserva las distancias relativas entre puntos. Los puntos que están cerca unos de los otros en el espacio original de entrada son mapeados a neuronas cercanas en SOM. Por esta razón, SOM es muy útil como herramienta de análisis de clases de datos de muchas dimensiones [Vesanto & Alhoniemi, 2000], y además tiene la capacidad de generalizar [Essenreiter et al., 1999], lo que implica que la red puede reconocer o caracterizar entradas que nunca antes ha encontrado.

K-means es un método iterativo que busca formar k clusters, con k predeterminado antes del inicio del proceso. K-means comienza particionando los datos en k subconjuntos no vacíos, calcula el centroide de cada partición como el punto medio del cluster y asigna cada dato al cluster cuyo centroide sea el más próximo. Luego

vuelve a particionar los datos iterativamente, hasta que no haya más datos que cambien de cluster de una iteración a la otra.

Otros algoritmos de clustering son K-medoids o PAM (Partition around medoids) y CLARA (Clustering LARge Applications) [Kaufman & Rousseeuw, 1990]. Este último permite manejar conjuntos de datos más grandes que el primero. CLARANS [Ng & Han, 1994] integra los algoritmos PAM y CLARA en uno.

Clasificación de datos

La clasificación se utiliza para clasificar un conjunto de datos basado en los valores de sus atributos. Por ejemplo, se podría clasificar a distintas personas para la otorgación de un préstamo en riesgo bajo, medio y alto, teniendo en cuenta información histórica de las mismas.

La clasificación encuentra las propiedades comunes entre un conjunto de objetos y los clasifica en diferentes clases, de acuerdo a un modelo de clasificación. Para construir este modelo, se utiliza un conjunto de entrenamiento, en el que cada instancia consiste en un conjunto de atributos y el valor de la clase a la cual pertenece. El objetivo de la clasificación es analizar los datos de entrenamiento y, mediante un método supervisado, desarrollar una descripción o un modelo para cada clase utilizando las

características disponibles en los datos. Esta descripción o modelo permite clasificar otras instancias, cuya clase es desconocida. El método se conoce como supervisado debido a que, para el conjunto de entrenamiento, se conoce la clase de pertenencia y se le indica al modelo si la clasificación que realiza es correcta o no. La construcción del modelo se realimenta de estas indicaciones del supervisor [Chen et al., 1996].

Los algoritmos mayormente utilizados para las tareas de clasificación son los algoritmos de inducción. En la actualidad existen numerosos enfoques de algoritmos de inducción y variedad en cada enfoque, el presente trabajo hará hincapié en aquellos orientados a generar árboles de decisión.

La clasificación basada en árboles de decisión es un método de aprendizaje supervisado que construye árboles de decisión a partir de un conjunto de entrenamiento.

Un sistema típico de construcción de árboles de decisión es ID3, que utiliza la teoría de la información para minimizar la cantidad de pruebas para clasificar un objeto. Al utilizar métodos heurísticos, ID3 garantiza un árbol simple, pero no necesariamente el más simple. Una extensión de ID3 es C4.5 [Quinlan, 1993a], que extiende el dominio de clasificación de atributos categóricos a numéricos. Un paso importante en la

construcción del árbol de decisión es la poda, la cual elimina las ramas no necesarias, resultando en una clasificación más rápida y una mejora en la precisión de la clasificación de datos [Han & Kamber, 2001].

Existen muchos otros algoritmos de clasificación de datos, incluyendo métodos estadísticos [Cheeseman & Stutz, 1996], como el análisis de regresión lineal [Elder IV & Pregibon, 1996]; algoritmos de machine learning [Cheeseman & Stutz, 1996]; redes neuronales [Lu et al., 1995], algoritmos genéticos y lógica difusa.

Asociación

La minería de reglas de asociación consiste en encontrar reglas de la forma $(A_1yA_2y...yA_m) \Rightarrow (B_1yB_2y...yB_n)$, donde A_i y B_j son valores de atributos del conjunto de datos [Chen et al., 1996]. Por ejemplo, se podría encontrar en un gran repositorio de datos de compras en un supermercado, la regla de asociación correspondiente a que si un cliente compra leche, entonces compra pan. Una regla de asociación es una sentencia probabilística acerca de la co-ocurrencia de ciertos eventos en una base de datos, y es particularmente aplicable a grandes conjuntos de datos [Hand et al., 2001].

Existen varios algoritmos que realizan el descubrimiento de reglas de asociación, uno de los más utilizados es Apriori.

CRISP-DM (Cross Industry Standard Process for Data Mining)

IBM (2012), considera que el CRISP-DM (Cross Industry Standard Process for Data Mining) proporciona una descripción normalizada del ciclo de vida de un proyecto estándar de análisis de datos, de forma análoga a como se hace en la ingeniería del software con los modelos de ciclo de vida de desarrollo de software. El modelo CRISP-DM cubre las fases de un proyecto, sus tareas respectivas, y las relaciones entre estas tareas. En este nivel de descripción no es posible identificar todas las relaciones; las relaciones podrían existir entre cualquier tarea según los objetivos, el contexto, y el interés del usuario sobre los datos.

La metodología CRISP-DM contempla el proceso de análisis de datos como un proyecto profesional, estableciendo así un contexto mucho más rico que influye en la elaboración de los modelos. Este contexto tiene en cuenta la existencia de un cliente que no es parte del equipo de desarrollo, así como el hecho de que el proyecto no sólo no acaba una vez se halla el modelo idóneo (ya que después se requiere un despliegue y un mantenimiento), sino que está relacionado con otros proyectos, y es preciso

documentarlo de forma exhaustiva para que otros equipos de desarrollo utilicen el conocimiento adquirido y trabajen a partir de él.

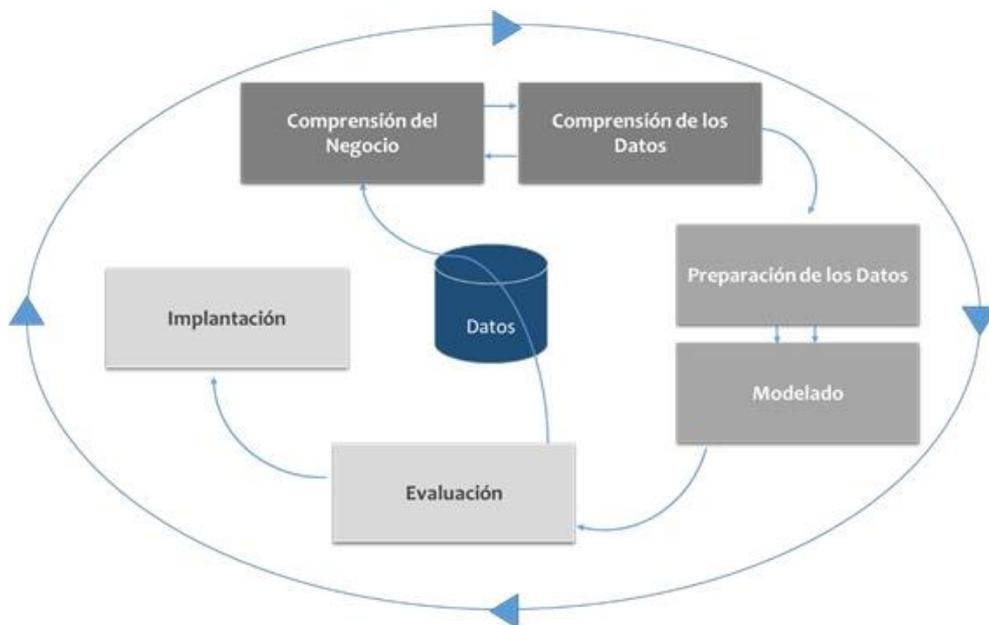


Figura 2. Modelo de proceso de la Metodología CRISP-DM ([CRISP-DM, 2000])

A continuación, se describen las fases del CRISP-DM.

1. Fase de comprensión del negocio o problema

La primera fase de la guía de referencia de la Metodología CRISP-DM, denominada fase de comprensión del negocio o problema (ver figura 3).

Probablemente la más importante, aglutina las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial o institucional, debido a que convierte en objetivos técnicos y en un plan de proyecto.

Sin lograr comprender dichos objetivos, ningún algoritmo por muy sofisticado que sea, permitirá obtener resultados fiables.

Por lo tanto, para obtener el mejor provecho de Data Mining, es necesario entender de la manera más completa el problema que se desea resolver, esto permitirá recolectar los datos correctos e interpretar correctamente los resultados.

Una descripción de cada una de las principales tareas que componen esta fase es la siguiente:

Determinar los objetivos del negocio.

Esta es la primera tarea a desarrollar y tiene como metas, determinar cuál es el problema que se desea resolver, por qué la necesidad de utilizar Data Mining y definir los criterios de éxito.

Los problemas pueden ser diversos como, por ejemplo, detectar fraude en el uso de tarjetas de crédito, detección de intentos de ingreso indebido a un sistema, asegurar el éxito de una determinada campaña publicitaria, entre otros.

En cuanto a los criterios de éxito, estos pueden ser de dos tipos:

Cualitativo, en cuyo caso un experto en el área de dominio, califica el resultado del proceso de DM.

Cuantitativo, por ejemplo, la respuesta de clientes ante la introducción de una nueva tarjeta de crédito.

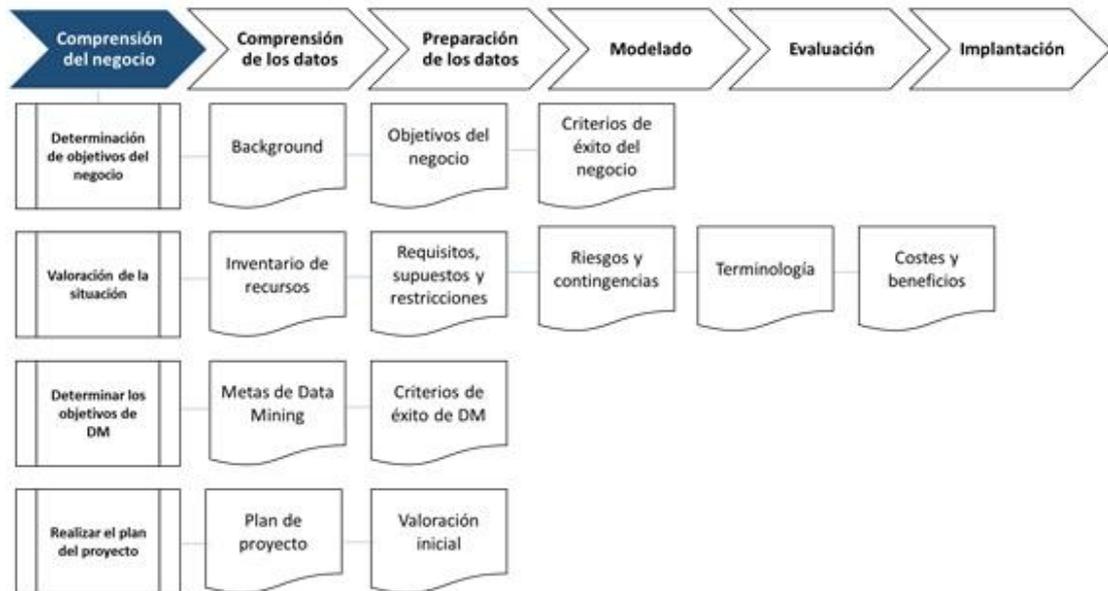


Figura 3. Fase de comprensión del negocio o problema

Evaluación de la situación.

En esta tarea se debe calificar el estado de la situación antes de iniciar el proceso de DM, considerando:

¿Cuál es el conocimiento previo disponible acerca del problema?

¿Se cuenta con la cantidad de datos requerida para resolver el problema?

¿Cuál es la relación coste beneficio de la aplicación de DM?

En esta fase se definen los requisitos del problema, tanto en términos de negocio como en términos de Data Mining.

Determinación de los objetivos de DM.

Esta tarea tiene como objetivo representar los objetivos del negocio en términos de las metas del proyecto de DM, como por ejemplo, si el objetivo del negocio es el desarrollo de una campaña publicitaria para incrementar la asignación de créditos hipotecarios, la meta de DM será por ejemplo, determinar el perfil de los clientes respecto de su capacidad de endeudamiento.

Producción de un plan del proyecto.

Finalmente, esta última tarea de la primera fase de CRISP-DM, tiene como meta desarrollar un plan para el proyecto, con los pasos a seguir y las técnicas a emplear.

2. Fase de comprensión de los datos

La segunda fase (figura 4), fase de comprensión de los datos, comprende la recolección inicial de datos, con el objetivo de establecer un primer contacto con el problema, familiarizándose con ellos, identificar su calidad y establecer las relaciones más evidentes que permitan definir las primeras hipótesis.

Esta fase junto a las próximas dos fases, son las que demandan el mayor esfuerzo y tiempo en un proyecto de DM. Por lo general si la organización cuenta con una base de datos corporativa, es deseable crear una nueva base de datos ad-hoc al proyecto de DM.

Debido a que durante el desarrollo del proyecto, probablemente se generen frecuentes y abundantes accesos a la base de datos a objeto de realizar consultas y probablemente modificaciones, lo cual podría generar muchos problemas.

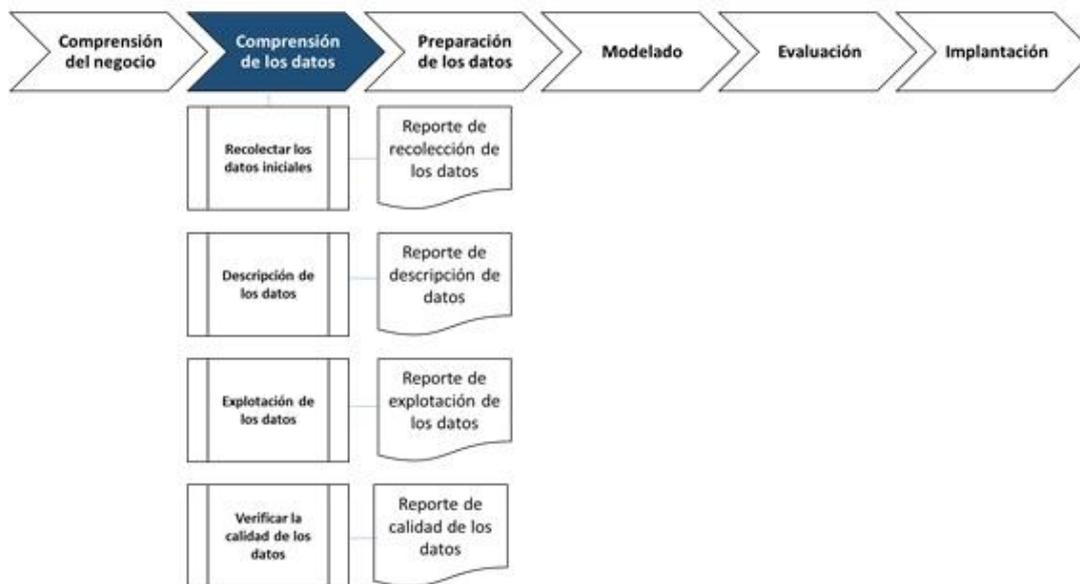


Figura 4. Fase de comprensión de los datos.

Las principales tareas a desarrollar en esta fase del proceso son:

Recolección de datos iniciales.

La primera tarea en esta segunda fase, es la recolección de los datos iniciales y su adecuación para el futuro procesamiento.

Esta tarea tiene como objetivo sobretodo:

Elaborar informes con una lista de los datos adquiridos

Su localización

Las técnicas utilizadas en su recolección

Los problemas y soluciones inherentes a este proceso.

Descripción de los datos.

Después de adquiridos los datos iniciales, estos deben ser descritos.

Este proceso involucra establecer volúmenes de datos, su identificación, el significado de cada campo y la descripción del formato inicial.

Exploración de datos.

A continuación, se procede a su exploración, cuyo fin es encontrar una estructura general para los datos.

Esto involucra la aplicación de pruebas estadísticas básicas, que revelen propiedades en los datos, se crean tablas de frecuencia y se construyen gráficos de distribución.

La salida de esta tarea es un informe de exploración de los datos.

Verificación de la calidad de los datos.

En esta tarea, se efectúan verificaciones sobre los datos, para determinar la consistencia de los valores individuales de los campos, la cantidad y distribución de los valores nulos, también para encontrar valores fuera de rango, los cuales pueden constituirse en ruido para el proceso.

En conclusión, la idea en este punto, es asegurar la completitud y corrección de los datos.

3. Fase de preparación de los datos

En esta fase y una vez efectuada la recolección inicial de datos, se procede a su preparación para adaptarlos a las técnicas de Data Mining que se utilicen posteriormente, tales como técnicas de visualización de datos, de búsqueda de relaciones entre variables u otras medidas para exploración de los datos.

La preparación de datos incluye las tareas generales de selección de datos a los que se va a aplicar una determinada técnica de modelado, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato.

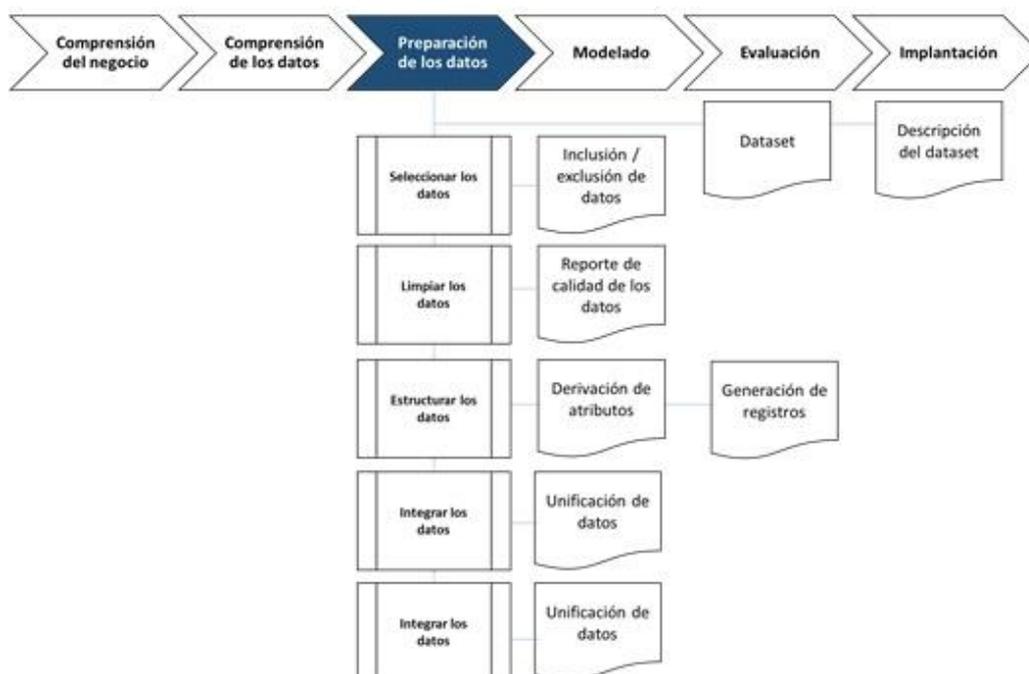


Figura 5. Fase de preparación de los datos.

Esta fase se encuentra relacionada con la fase de modelado, puesto que en función de la técnica de modelado elegida, los datos requieren ser procesados de diferentes

formas. Es así que las fases de preparación y modelado interactúan de forma permanente.

La figura 5, ilustra las áreas de que se compone ésta, e identifica sus salidas. Una descripción de las tareas involucradas en esta fase es la siguiente:

Selección de datos

En esta etapa, se selecciona un subconjunto de los datos adquiridos en la fase anterior, apoyándose en criterios previamente establecidos en las fases anteriores:

Calidad de los datos en cuanto a completitud

Corrección de los datos y limitaciones en el volumen o en los tipos de datos que están relacionadas con las técnicas de DM seleccionadas.

Limpieza de los datos

Esta tarea complementa a la anterior, y es una de las que más tiempo y esfuerzo consume, debido a la diversidad de técnicas que pueden aplicarse para optimizar la calidad de los datos a objeto de prepararlos para la fase de modelación.

Algunas de las técnicas a utilizar para este propósito son:

Normalización de los datos

Discretización de campos numéricos

Tratamiento de valores ausentes

Reducción del volumen de datos

Estructuración de los datos.

Esta tarea incluye las operaciones de preparación de los datos tales como:

Generación de nuevos atributos a partir de atributos ya existentes

Integración de nuevos registros o transformación de valores para atributos existentes.

Integración de los datos.

La integración de los datos, involucra la creación de nuevas estructuras, a partir de los datos seleccionados, por ejemplo:

Generación de nuevos campos a partir de otros existentes

Creación de nuevos registros

Fusión de tablas campos o nuevas tablas donde se resumen características de múltiples registros o de otros campos en nuevas tablas de resumen.

Formateo de los datos

Esta tarea consiste principalmente, en la realización de transformaciones sintácticas de los datos sin modificar su significado, esto, con la idea de permitir o facilitar el empleo de alguna técnica de DM en particular, como por ejemplo:

La reordenación de los campos y/o registros de la tabla o el ajuste de los valores de los campos a las limitaciones de las herramientas de modelación:

Eliminar comas, tabuladores, caracteres especiales, máximos y mínimos para las cadenas de caracteres, entre otros.

4. Fase de modelado

En esta fase de CRISP-DM, se seleccionan las técnicas de modelado más apropiadas para el proyecto de Data Mining específico.

Las técnicas a utilizar en esta fase se eligen en función de los siguientes criterios:

- Ser apropiada al problema
- Disponer de datos adecuados
- Cumplir los requisitos del problema
- Tiempo adecuado para obtener un modelo

– Conocimiento de la técnica

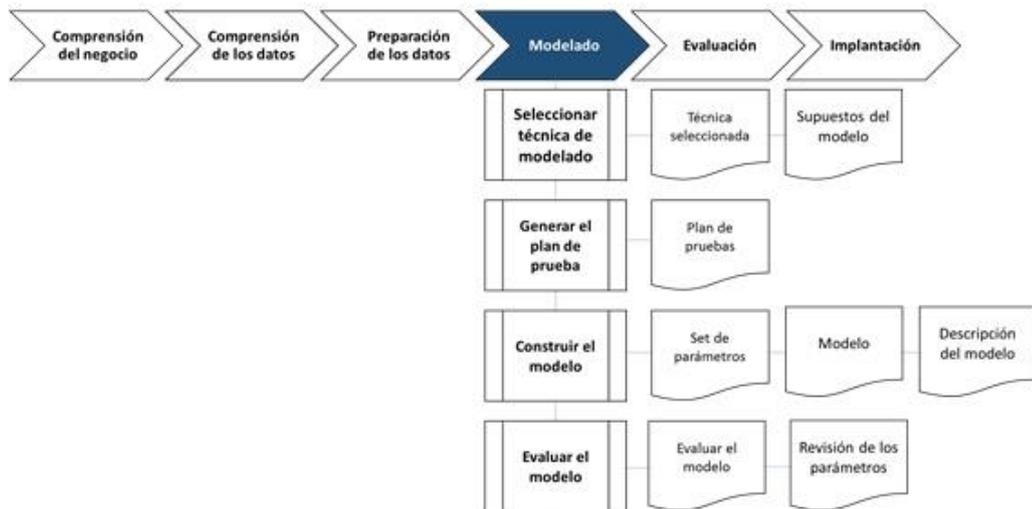


Figura 6. Fase del modelado.

Previamente al modelado de los datos, se debe determinar un método de evaluación de los modelos que permita establecer el grado de bondad de ellos.

Después de concluir estas tareas genéricas, se procede a la generación y evaluación del modelo.

Los parámetros utilizados en la generación del modelo, dependen de las características de los datos y de las características de precisión que se quieran lograr con el modelo.

La figura 6 ilustra las tareas y resultados que se obtienen en esta fase. Una descripción de las principales tareas de esta fase es la siguiente:

Selección de la técnica de modelado.

Esta tarea consiste en la selección de la técnica de DM más apropiada al tipo de problema a resolver.

Para esta selección, se debe considerar el objetivo principal del proyecto y la relación con las herramientas de DM existentes.

Por ejemplo, si el problema es de clasificación, se podrá elegir de entre árboles de decisión, k-nearest neighbour o razonamiento basado en casos (CBR); si el problema es de predicción, análisis de regresión, redes neuronales; o si el problema es de segmentación, redes neuronales, técnicas de visualización, etc.

Generación del plan de prueba.

Una vez construido un modelo, se debe generar un procedimiento destinado a probar la calidad y validez del mismo.

Por ejemplo, en una tarea supervisada de DM como la clasificación, es común usar la razón de error como medida de la calidad.

Entonces, típicamente se separan los datos en dos conjuntos, uno de entrenamiento y otro de prueba, para luego construir el modelo basado en el conjunto de entrenamiento y medir la calidad del modelo generado con el conjunto de prueba.

Construcción del Modelo

Después de seleccionada la técnica, se ejecuta sobre los datos previamente preparados para generar uno o más modelos. Todas las técnicas de modelado tienen un conjunto de parámetros que determinan las características del modelo a generar. La selección de los mejores parámetros es un proceso iterativo y se basa exclusivamente en los resultados generados. Estos deben ser interpretados y su rendimiento justificado.

Evaluación del modelo

En esta tarea, los ingenieros de DM interpretan los modelos de acuerdo al conocimiento preexistente del dominio y los criterios de éxito preestablecidos. Expertos en el dominio del problema juzgan los modelos dentro del contexto del dominio y expertos en Data Mining aplican sus propios criterios (seguridad del conjunto de prueba, pérdida o ganancia de tablas, entre otros).

5. Fase de evaluación

En esta fase se evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema.

Debe considerarse además, que la fiabilidad calculada para el modelo se aplica solamente para los datos sobre los que se realizó el análisis.

Es preciso revisar el proceso, teniendo en cuenta los resultados obtenidos, para poder repetir algún paso anterior, en el que se haya posiblemente cometido algún error.

Considerar que se pueden emplear múltiples herramientas para la interpretación de los resultados.

Las matrices de confusión Edelstein, 1999 son muy empleadas en problemas de clasificación y consisten en una tabla que indica cuantas clasificaciones se han hecho para cada tipo, la diagonal de la tabla representa las clasificaciones correctas.

Si el modelo generado es válido en función de los criterios de éxito establecidos en la fase anterior, se procede a la explotación del modelo. La figura N°7 Detalla las tareas que componen esta fase y los resultados que se deben obtener.

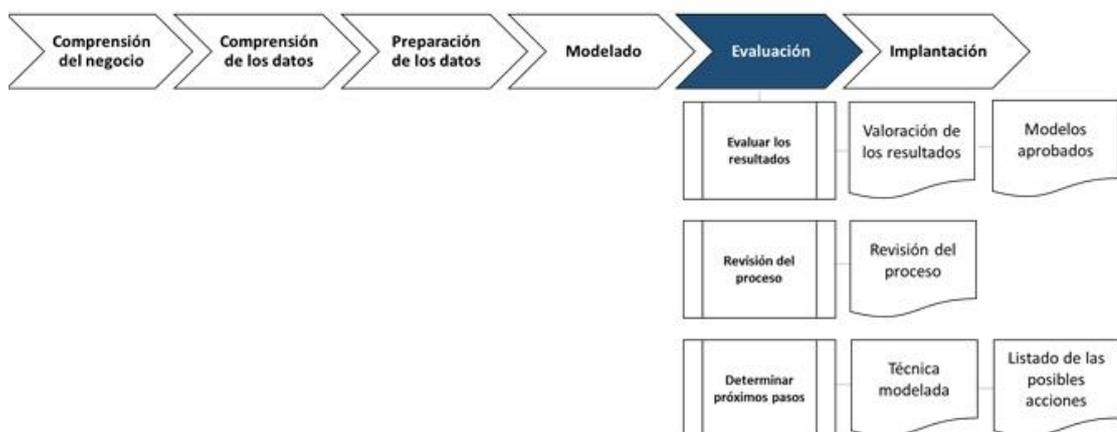


Figura 7. Fase de evaluación ([CRISP-DM, 2000]).

Las tareas involucradas en esta fase del proceso son las siguientes:

Evaluación de los resultados

En los pasos de evaluación anteriores, se trataron factores tales como la exactitud y generalidad del modelo generado.

Esta tarea involucra la evaluación del modelo en relación a los objetivos del negocio y busca determinar si hay alguna razón de negocio para la cual, el modelo sea deficiente, o si es aconsejable probar el modelo, en un problema real si el tiempo y restricciones lo permiten.

Además de los resultados directamente relacionados con el objetivo del proyecto

¿Es aconsejable evaluar el modelo en relación a otros objetivos distintos a los originales?

Esto podría revelar información adicional.

Revisión del proceso

El proceso de revisión, se refiere a calificar al proceso entero de DM, a objeto de identificar elementos que pudieran ser mejorados.

Determinación de futuras fases

Si se ha determinado que las fases hasta este momento han generado resultados satisfactorios, podría pasarse a la fase siguiente, en caso contrario podría decidirse por otra iteración desde la fase de preparación de datos o de modelación con otros parámetros.

Podría ser incluso que en esta fase se decida partir desde cero con un nuevo proyecto de DM.

6. Fase de implementación

En esta fase (figura 8), y una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio, ya sea que el analista recomiende acciones basadas en la observación del modelo y sus resultados, ya sea aplicando el modelo a diferentes conjuntos de datos o como parte del proceso.

Por ejemplo, en análisis de riesgo crediticio, detección de fraudes, etc.

Generalmente un proyecto de Data Mining no concluye en la implantación del modelo, pues se deben documentar y presentar los resultados de manera comprensible para el usuario, con el objetivo de lograr un incremento del conocimiento.

Por otra parte, en la fase de explotación se debe asegurar el mantenimiento de la aplicación y la posible difusión de los resultados.

Las tareas que se ejecutan en esta fase son las siguientes:

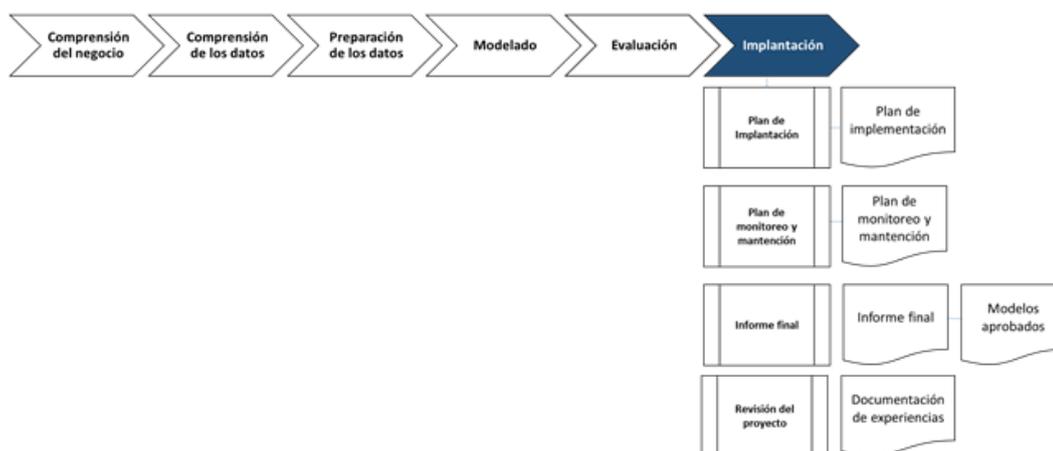


Figura 8. Fase de implementación ([CRISP-DM, 2000])

Plan de implementación

Para implementar el resultado de DM en la organización, esta tarea toma los resultados de la evaluación y concluye una estrategia para su implementación.

Si un procedimiento general se ha identificado para crear el modelo, este procedimiento debe ser documentado para su posterior implementación.

Monitorización y Mantenimiento

Si los modelos resultantes del proceso de Data Mining son implementados en el dominio del problema como parte de la rutina diaria, es aconsejable preparar estrategias de monitorización y mantenimiento para ser aplicadas sobre los modelos.

La retroalimentación generada por la monitorización y mantenimiento pueden indicar si el modelo está siendo utilizado apropiadamente

Informe Final

Es la conclusión del proyecto de DM realizado.

Dependiendo del plan de implementación, este informe puede ser sólo un resumen de los puntos importantes del proyecto y la experiencia lograda o puede ser una presentación final que incluya y explique los resultados logrados con el proyecto.

Revisión del proyecto

En este punto se evalúa qué fue lo correcto y qué lo incorrecto, qué es lo que se hizo bien y qué es lo que se requiere mejorar

Árboles de Decisión

Un árbol de decisión es un modelo de predicción utilizado en diversos ámbitos que van desde la inteligencia artificial hasta la Economía. Dado un conjunto de datos se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema (Lior Rokach y Oded Maimon, 2008)

Elementos

Los árboles de decisión están formados por nodos, vectores de números, flechas y etiquetas.

Nodo: se puede definir como el momento en el que se ha de tomar una decisión de entre varias posibles, lo que va haciendo que a medida que aumenta el número de nodos aumente el número de posibles finales a los que puede llegar el individuo. Esto hace que un árbol con muchos nodos sea complicado de dibujar a mano y de analizar debido a la existencia de numerosos caminos que se pueden seguir.

Los vectores de números serían la solución final a la que se llega en función de las diversas posibilidades que se tienen, dan las utilidades en esa solución.

Las flechas son las uniones entre un nodo y otro y representan cada acción distinta.

Las etiquetas se encuentran en cada nodo y cada flecha y dan nombre a cada acción.

Análisis Factorial

El Análisis Factorial es un método multivariante que pretende expresar p variables observables como una combinación lineal de m variables hipotéticas o latentes, denominadas factores. Tiene una formulación parecida al Análisis de Componentes Principales, pero el modelo que relaciona variables y factores es diferente en Análisis Factorial. Si la matriz de correlaciones existe, las componentes principales también existen, mientras que el modelo factorial podría ser aceptado o no mediante un test estadístico. (Lior Rokach y Oded Maimon, 2008)

Análisis Factorial Confirmatorio

Esta técnica, permite contrastar un modelo construido con antelación, en el que el investigador establece a priori el conjunto total de las relaciones entre los elementos que lo configuran, en el factorial confirmatorio se supone que el investigador es capaz de aventurar a priori la estructura de los datos -preferiblemente en función de una teoría bien establecida- y sólo precisa confirmar que esa estructura puede también obtenerse empíricamente. persiguen un objetivo: “explicar las covarianzas o correlaciones entre un conjunto de variables observadas o medidas a través de un conjunto reducido de variables latentes o factores” (Bollen, 1989, p. 226).

2.2.2. Definición de términos

ENAH: Encuesta Nacional de Hogares

Estandarización: Es un procedimiento que facilita la comparación de dos variables en una misma escala de medición con media cero y varianza uno.

CHAID: detección automática de interacciones mediante chi-cuadrado.

Aprendizaje supervisado: Los datos del conjunto de aprendizaje tienen dos tipos de atributos: los datos propiamente dichos y cierta información relativa a la solución del problema. (Viñuela y León, 2004)

Aprendizaje no supervisado: Los datos del conjunto de aprendizaje sólo tienen información de los ejemplos, y no hay nada que permita guiar en el proceso de aprendizaje. En este caso, no existe profesor externo que determine el aprendizaje. (Viñuela y León, 2004)

Calidad del servicio: Está constituida por todos los atributos o propiedades que lo conforman y que le otorgan valor. La calidad de los productos puede ser fácilmente medible pero no así la calidad de los servicios debido a sus características de intangibilidad, inseparabilidad, heterogeneidad y caducidad (Zeithaml y Parasuraman, 2004).

Servicio educativo: Son los órganos de apoyo permanente a la labor docente del profesorado y de los centros docentes en particular (Vera, J., 2014).

**CAPÍTULO III: HIPÓTESIS Y OPERACIONALIZACIÓN
DE VARIABLES**

3.1. Hipótesis

La calidad del servicio educativo peruano se relaciona con la lectoescritura en estudiantes de educación primaria del Módulo Educación de la ENAHO trimestre 2, 2016, mediante la Minería de datos de manera directa.

3.2. Operacionalización de variables

Las variables de la presente investigación son:

<i>Variable</i>	<i>Indicadores</i>	<i>Tipo de variables</i>
Variable independiente: Calidad del servicio educativo	Infraestructura	Cualitativo
	Equipamiento	Cualitativo
	Enseñanza docente	Cualitativo
	Materiales	Cualitativo
	Apoyo de padres	Cualitativo
Variable dependiente Lectoescritura	Lectura	Cualitativo
	Escritura	Cualitativo

CAPÍTULO IV: METODOLOGÍA

4.1. Materiales y lugar de ejecución

4.1.1. Lugar

Microdatos del Instituto Nacional de Estadística e Informática (INEI) ubicado en la página web: <http://iinei.inei.gob.pe/microdatos/>.

4.1.2. Materiales

Los materiales utilizados para el presente trabajo de tesis fueron los útiles de escritorio como papel bond, lapiceros, fichas de recolección de datos, USB, cuadernos de apunte.

4.1.3. Equipos

Los equipos utilizados en el presente trabajo de tesis fueron una laptop, una impresora, fotocopidora.

4.1.4. Recurso humano

El presente trabajo de tesis cuenta con la asesoría de un docente de la escuela académico profesional de Estadística e Informática y dos investigadores quienes son los autores de la presente investigación.

4.2. Métodos

4.2.1. Técnicas de recolección de datos

Para la presente investigación se recopiló información del sistema de base de datos del INEI (<http://inei.inei.gob.pe/microdatos/>).

4.2.2. Tipo de estudio

- *Según su propósito o finalidad:* Aplicada
- *Según su alcance temporal:* Transversal
- *Según su enfoque:* Cuantitativo
- *Según su nivel de conocimientos que se adquieren o profundidad:* correlacional
- *Según el método utilizado:* Analítica, inductiva.

4.2.3. Diseño de la investigación

La presente investigación es de diseño no experimental, de tipo descriptivo correlacional.

4.2.3. Población y muestra

4.2.3.1. Población: Ocupantes residentes de las viviendas particulares del área urbana y rural del país que son estudiantes del nivel primaria.

4.2.3.2. Muestra: Conformada por 29585 encuestas de la ENAHO del segundo trimestre del año 2016. Además, fue constituida por las encuestas que satisfagan las características de estudio, de limpieza y selección producto de la minería de datos; es decir, que los encuestados sean estudiantes de educación primaria.

4.2.4. Instrumentos de recopilación de datos

- Técnica:

- Documentaria: Se recopiló a partir de las encuestas de la ENAHO del segundo trimestre del año 2016.

- Instrumento:

- Encuesta nacional de hogares del INEI.

4.2.5. Análisis Estadístico e interpretación de la información

Para el procesamiento y organización de los datos se utilizó el software SPSS v24 y el R-Project v3.4.1.

El análisis e interpretación estadística de los datos para la presente investigación se realizó por medio de las medidas de tendencia central y dispersión, presentados en gráficos estadísticos y organizados en los cuadros estadísticos. Además, se hizo uso del análisis de la minería de datos.

CAPÍTULO V: ANÁLISIS Y DISCUSIÓN DE RESULTADOS

5.1. Comprensión del problema

Uno de los problemas a solucionar fue la fiabilidad del instrumento de medición de la calidad del servicio educativo peruano desarrollado por el INEI, como parte de la encuesta nacional de hogares – ENAHO, así como, el perfil de dichos indicadores de calidad asociados a la lectoescritura de los estudiantes de educación primaria.

5.2. Comprensión de los datos

Los datos fueron recolectados de la página web del INEI, del repositorio de base de datos, del enlace microdatos (<http://inei.inei.gob.pe/microdatos/>), de la base de datos de la ENAHO metodología actualizada, de las condiciones de vida y pobreza, del año 2016 del trimestre 3, contando con un total de 8 módulos:

- Características de la Vivienda y del Hogar
- Características de los Miembros del Hogar
- Educación
- Salud
- Empleo e ingresos
- Gastos en Alimentos y Bebidas (Módulo 601)
- Programas Sociales (Miembros del Hogar)
- Gobernabilidad, Democracia y Transparencia

De los cuales, para la minería de datos se utilizó el módulo de:

- **Características de los Miembros del Hogar:** compuesta por un archivo de 39 campos de variables y 35283 registros de datos.
- **Educación:** compuesta por dos archivos, donde el primer archivo contiene 455 campos de variables y 31759 registros de datos. El segundo archivo contiene 29 campos de variables y 675 registros.

Las variables seleccionadas son:

- Sabe leer y escribir
- Calidad del servicio educativo en infraestructura
- Calidad del servicio educativo en equipamiento
- Calidad del servicio educativo en enseñanza de los maestros
- Calidad del servicio educativo en dotación de materiales educativos
- Calidad del servicio educativo en apoyo a la participación de los padres de familias
- Idioma que habla
- Tipo de centro de estudios
- Resultados académicos del año pasado
- Grado matriculado
- Uso de internet
- Sexo del estudiante
- Edad del estudiante
- Región donde habita

5.3. Preparación de los datos

En la preparación de los datos se fusionaron los tres archivos anteriormente mencionados, para su posterior filtro y selección de las variables estudiadas. Se eliminaron los registros con campos en blanco, así como, personas mayores de edad que no correspondían a los estudiantes de educación básica regular de nivel primaria. Después de la limpieza de los datos, la base de datos estuvo conformado por 15 campos antes señalados y un total de 3083 casos.

Para el objetivo de validación de la calidad del servicio educativo peruano, mediante el análisis factorial confirmatorio por medio del programa AMOS se requirió las siguientes variables formateadas a la misma escala de medición de la ENAHO 2016:

- Infraestructura (paredes, techos, pisos)
- Equipamiento (Carpetas, Laboratorios, Computadoras, etc.)
- Enseñanza de los maestros
- Dotación de materiales educativos (Libros, Láminas, maquetas, etc.)
- Apoyo a la participación de los padres de familias (Promueven la participación de los padres de familia, escuela de padre)

Mientras que para los árboles de decisión se hicieron uso de las variables en su forma original observada mediante la ENAHO 2016:

- Sabe leer y escribir
- Idioma o lengua materna que aprendió en su niñez

- El resultado que obtuvo en el 2015
- Grado de Estudios en el que está matriculado
- Sexo del estudiante
- Edad en años cumplidos del estudiante
- Región donde habita el estudiante
- Tipo de Centro de Estudios al que asistió en el 2015
- Uso del servicio de Internet por cabina pública

5.4. Modelado

La minería de los datos de estudio se realizó por medio del paquete R-Project v3.4.3 y RStudio v1.1.423, previamente los datos originales se encontraron almacenados en el SPSS v24, para ello los registros fueron limpiados, estandarizados y exportados al R-Project.

El código de importación de la base de datos en R-Project a partir de una base de datos almacenada en SPSS es:

```
### Código de importación de base de datos de SPSS a R ###  
library(haven)  
BaseDatos <- read_sav("D:/BaseDatos.sav")  
View(BaseDatos)  
### Fin de importación de base de datos ###  
|
```

Dando como resultado una base de datos con 15 variables y 3083 observaciones.

AÑO	MES	CONGLOME	VIVIENDA	HOGAR	CODPERSO	UBIGEO	ubigeo2
Año	Mes de Ejecución de la Encuesta	N° de Conglomerado	N° de Selección de la Vivienda	Hogar	Código de la persona	Código de Distrito (Ubicación Geográfica)	
1	2016	04	005080	221	11	03	10401
2	2016	04	005080	222	11	04	10401
3	2016	06	005439	115	11	03	30102
4	2016	06	005514	002	11	02	40102
5	2016	05	005828	053	11	05	50612
6	2016	06	006293	083	11	04	81303
7	2016	05	006367	118	11	03	90101
8	2016	05	006629	126	11	03	100109
9	2016	05	006831	120	11	03	110301
10	2016	04	008067	041	11	02	150601
11	2016	06	008358	080	11	07	160210
12	2016	04	010162	132	11	03	120699
13	2016	06	010400	031	11	02	10101

Del total de variables incluidas en la base de datos, sólo haremos uso a aquellas relacionadas con los objetivos de estudio, tales como:

- 302. ¿Sabe leer y escribir?
- 308B1. Infraestructura (paredes, techos, pisos)
- 308B2. Equipamiento (Carpetas, Laboratorios, Computadoras, etc.)
- 308B3. Enseñanza de los maestros
- 308B4. Dotación de materiales educativos (Libros, Láminas, maquetas, etc.)
- 308B5. Apoyo a la participación de los padres de familia (Promueven la participación de los padres de familia, escuela de padres, etc.)

El código de asignación de etiqueta a la variable ¿Sabe leer y escribir? Con respuestas “No” y “si”:

```
### Asignación de etiqueta a la variable ¿Sabe leer y escribir? ###
LeeEscribe_f<-factor(BaseDatos$P302)
levels(LeeEscribe_f)<-c("Si", "No")
LeeEscribe_f
### Fin de asignación de etiqueta a la variable ¿Sabe leer y escribir? ###
```

La distribución de datos de la ¿Sabe leer y escribir? se presenta a continuación:

```
### Creación de tablas de frecuencias ###
tLeeEscribe<-table(LeeEscribe_f)
tLeeEscribe
### Fin de creación de tabla de distribución de frecuencias ###
```

Tabla 1

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria del Módulo Educación de la ENAHO trimestre 2, 2016

¿Sabe leer y escribir?	n	%
No	539	17%
Si	2544	83%
Total	3083	100%

Fuente: Base de datos

De la muestra de estudio se observa que el 83% de los estudiantes de educación primaria si saben leer y escribir, mientras que el 17% de los restantes no saben leer ni escribir.

Tabla 2

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según idioma o lengua materna que aprendió en su niñez, del Módulo Educación de la ENAHO trimestre 2, 2016

Idioma o lengua materna que aprendió en su niñez	Sabe leer y escribir			
	Si		No	
	n	%	n	%
Quechua	281	11%	41	8%
Aymará	9	0%	0	0%
Otra lengua nativa	96	4%	56	10%
Castellano	2147	84%	432	80%
Inglés	0	0%	0	0%
Portugués	9	0%	3	1%

Otra lengua extranjera	2	0%	0	0%
Es sordo mudo	0	0%	7	1%
Total	2544	100%	539	100%

Fuente: Base de datos

De la tabla 2, se observa que el 84% de los estudiantes de educación primaria que saben leer y escribir aprendieron en su niñez el castellano como idioma o lengua materna, mientras que el 80% de los estudiantes de educación primaria que no saben leer y ni escribir aprendieron en su niñez el castellano como idioma o lengua materna.

Tabla 3

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según centro de estudios del Módulo Educación de la ENAHO trimestre 2, 2016

Centro de estudios	Sabe leer y escribir			
	Si		No	
	n	%	n	%
Estatad	2522	99%	529	98%
No estatal	22	1%	10	2%
Total	2544	100%	539	100%

Fuente: Base de datos

De la tabla 3, se observa que el 99% de los estudiantes de educación primaria que saben leer y escribir estudian en un centro de estudios estatal, mientras que el 98% de los estudiantes de educación primaria que no saben leer y ni escribir estudian en un centro de estudio no estatal.

Tabla 4

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según resultado que obtuvo el año pasado del Módulo Educación de la ENAHO trimestre 2, 2016

Resultado que obtuvo el año pasado	Sabe leer y escribir			
	Si		No	
	n	%	n	%
Aprobado	2392	94%	164	30%
Desaprobado	55	2%	19	4%
Retirado	2	0%	11	2%
No aprueba ni desaprueba	95	4%	345	64%
Total	2544	100%	539	100%

Fuente: Base de datos

De la tabla 4, se observa que el 94% de los estudiantes de educación primaria que saben leer y escribir aprobaron sus estudios el año pasado, mientras que el 30% de los estudiantes de educación primaria que no saben leer y ni escribir aprobaron sus estudios el año pasado.

Tabla 5

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según resultado que obtuvo el año pasado del Módulo Educación de la ENAHO trimestre 2, 2016

¿Cuál es el Año o Grado de Estudios en el que está matriculado?	Sabe leer y escribir			
	Si		No	
	n	%	n	%
1ro	96	4%	354	66%
2do	406	16%	123	23%
3ro	477	19%	40	7%

4to	503	20%	16	3%
5to	512	20%	6	1%
6to	550	22%	0	0%
Total	2544	100%	539	100%

Fuente: Base de datos

De la tabla 5, se observa que el 22% de los estudiantes de educación primaria que saben leer y escribir se encuentran matriculados en el 6to grado, mientras que el 66% de los estudiantes de educación primaria que no saben leer y ni escribir se encuentran matriculados en el primer grado.

Tabla 6

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según uso del servicio de internet del Módulo Educación de la ENAHO trimestre 2, 2016

¿hizo uso del servicio de internet?	Sabe leer y escribir			
	Si		No	
	n	%	n	%
Si	519	20%	28	5%
No	2025	80%	511	95%
Total	2544	100%	539	100%

Fuente: Base de datos

De la tabla 6, se observa que el 20% de los estudiantes de educación primaria que saben leer y escribir si hicieron uso del servicio de internet, mientras que el 5% de los estudiantes de educación primaria que no saben leer y ni escribir no hicieron uso del servicio de internet.

Tabla 7

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según sexo en el Módulo Educación de la ENAHO trimestre 2, 2016

Sexo	Sabe leer y escribir			
	Si		No	
	n	%	n	%
Hombre	1306	51%	282	52%
Mujer	1238	49%	257	48%
Total	2544	100%	539	100%

Fuente: Base de datos

De la tabla 7, se observa que el 51% de los estudiantes de educación primaria que saben leer y escribir son varones, mientras que el 52% de los estudiantes de educación primaria que no saben leer y ni escribir son varones.

Tabla 8

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según edad en el Módulo Educación de la ENAHO trimestre 2, 2016

Edad	Sabe leer y escribir			
	Si		No	
	n	%	n	%
5 años	2	0%	2	0%
6 años	81	3%	301	56%
7 años	345	14%	132	24%
8 años	441	17%	45	8%
9 años	474	19%	28	5%
10 años	512	20%	14	3%
11 años	463	18%	4	1%

12 años	134	5%	5	1%
13 años	54	2%	0	0%
14 años	18	1%	2	0%
15 años	20	1%	6	1%
Total	2544	100%	539	100%

Fuente: Base de datos

De la tabla 8, se observa que el 20% de los estudiantes de educación primaria que saben leer y escribir tienen 10 años de edad, mientras que el 56% de los estudiantes de educación primaria que no saben leer y ni escribir tienen 6 años de edad.

Tabla 9

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según región en el Módulo Educación de la ENAHO trimestre 2, 2016

Región	Sabe leer y escribir			
	Si		No	
	n	%	n	%
Costa	774	30%	124	23%
Sierra	1066	42%	186	35%
Selva	704	28%	229	42%
Total	2544	100%	539	100%

Fuente: Base de datos

De la tabla 9, se observa que el 42% de los estudiantes de educación primaria que saben leer y escribir provienen de la región sierra, mientras que el 42% de los estudiantes de educación primaria que no saben leer y ni escribir provienen de la región selva.

Tabla 10

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según servicio educativo: infraestructura en el Módulo Educación de la ENAHO trimestre 2, 2016

Infraestructura (paredes, techos, pisos)	Sabe leer y escribir				Total	
	Si		No			
	n	%	n	%	n	%
Muy malo	25	1%	6	1%	31	1%
Malo	367	14%	83	15%	450	15%
Bueno	2057	81%	434	81%	2491	81%
Muy bueno	95	4%	16	3%	111	4%
Total	2544	100%	539	100%	3083	100%

Fuente: Base de datos

De la tabla 10, se observa que el 81% de los estudiantes de educación primaria que saben leer y escribir califican el servicio educativo en la infraestructura como buena, mientras que el 81% de los estudiantes de educación primaria que no saben leer y ni escribir califican el servicio educativo en la infraestructura como buena.

Tabla 11

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según servicio educativo: equipamiento en el Módulo Educación de la ENAHO trimestre 2, 2016

Equipamiento (Carpetas, Laboratorios, Computadoras, etc.)	Sabe leer y escribir				Total	
	Si		No			
	n	%	n	%	n	%
Muy malo	18	1%	5	1%	23	1%
Malo	493	19%	95	18%	588	19%
Bueno	1964	77%	427	79%	2391	78%
Muy bueno	69	3%	12	2%	81	3%
Total	2544	100%	539	100%	3083	100%

Fuente: Base de datos

De la tabla 11, se observa que el 77% de los estudiantes de educación primaria que saben leer y escribir califican el servicio educativo en el equipamiento como bueno, mientras que el 79% de los estudiantes de educación primaria que no saben leer y ni escribir califican el servicio educativo en el equipamiento como bueno.

Tabla 12

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según servicio educativo: enseñanza de los maestros en el Módulo Educación de la ENAHO trimestre 2, 2016

Enseñanza de los maestros	Sabe leer y escribir				Total	
	Si		No		n	%
	n	%	n	%		
Muy malo	14	1%	4	1%	18	1%
Malo	226	9%	45	8%	271	9%
Bueno	2166	85%	466	86%	2632	85%
Muy bueno	138	5%	24	4%	162	5%
Total	2544	100%	539	100%	3083	100%

Fuente: Base de datos

De la tabla 12, se observa que el 85% de los estudiantes de educación primaria que saben leer y escribir califican el servicio educativo en la enseñanza de los maestros como bueno, mientras que el 86% de los estudiantes de educación primaria que no saben leer y ni escribir califican el servicio educativo en la enseñanza de los maestros como bueno.

Tabla 13

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según servicio educativo: dotación de materiales educativos en el Módulo Educación de la ENAHO trimestre 2, 2016

Dotación de materiales educativos (Libros, Láminas, maquetas, etc.)	Sabe leer y escribir				Total	
	Si		No		n	%
	n	%	n	%		
Muy malo	8	0%	1	0%	9	0%
Malo	219	9%	48	9%	267	9%
Bueno	2231	88%	472	88%	2703	88%
Muy bueno	86	3%	18	3%	104	3%
Total	2544	100%	539	100%	3083	100%

Fuente: Base de datos

De la tabla 13, se observa que el 88% de los estudiantes de educación primaria que saben leer y escribir califican el servicio educativo en la dotación de materiales educativos como buena, mientras que el 88% de los estudiantes de educación primaria que no saben leer y ni escribir califican el servicio educativo en la dotación de materiales educativos como buena.

Tabla 14

¿Sabe leer y escribir? respuestas de los estudiantes de educación primaria según servicio educativo: Apoyo a la participación de los padres de familia en el Módulo Educación de la ENAHO trimestre 2, 2016

Apoyo a la participación de los padres de familia (Promueven la participación de los padres de familia, escuela de padres, etc)	Sabe leer y escribir				Total	
	Si		No			
	n	%	n	%	n	%
Muy malo	10	0%	0	0%	10	0%
Malo	286	11%	60	11%	346	11%
Bueno	2168	85%	468	87%	2636	86%
Muy bueno	80	3%	11	2%	91	3%
Total	2544	100%	539	100%	3083	100%

Fuente: Base de datos

De la tabla 14, se observa que el 85% de los estudiantes de educación primaria que saben leer y escribir califican el servicio educativo en el apoyo a la participación de los padres de familia como bueno, mientras que el 87% de los estudiantes de educación primaria que no saben leer y ni escribir califican el servicio educativo en el apoyo a la participación de los padres de familia como bueno.

Análisis factorial confirmatorio respecto a la calidad del servicio educativo

El presente análisis es con la finalidad de confirmar que los cinco ítems medidos expliquen la calidad del servicio educativo, para ello se tiene identificado las siguientes características iniciales del modelo:

Variables observadas:

- Infraestructura (P308B1)

- Equipamiento (P308B2)
- Enseñanza del docente (P308B3)
- Dotación de materiales educativos (P308B4)
- Apoyo a la participación de padres de familia (P308B5)

Variable latente: factor común

- Servicio educativo (ξ).

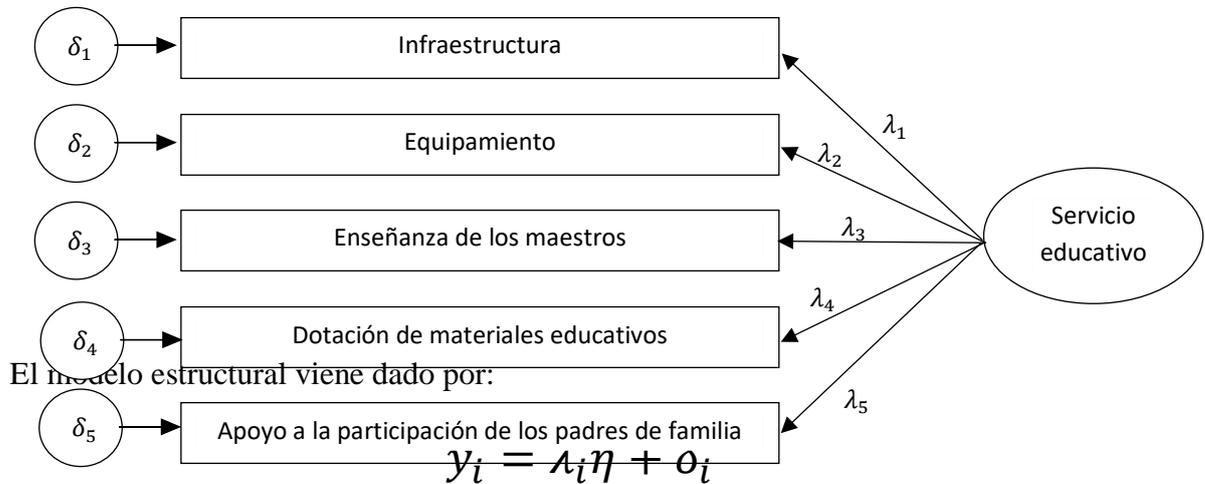
Variable latente: factor específico o errores

- Factor específico de infraestructura (δ_1).
- Factor específico de equipamiento (δ_2).
- Factor específico de enseñanza docente (δ_3).
- Factor específico de dotación de materiales educativos (δ_4).
- Factor específico de apoyo a la participación de padres de familia (δ_5).

Intensidad de relación (carga factorial) entre el servicio educativo y:

- La infraestructura (λ_1).
- El equipamiento (λ_2).
- La enseñanza docente (λ_3).
- La dotación de materiales educativos (λ_4).
- La participación de padres de familia (λ_5).

Modelo de Análisis factorial confirmatorio:



Para la estimación del modelo estructural, se hizo necesario comprobar el supuesto de normalidad multivariante, para ello se probó mediante la normalidad multivariante de Mardia respecto a la Kurtosis:

$$[kurt - k(k + 2)] \sqrt{\frac{n}{8k(k + 2)}} \sim N(0,1)$$

Donde:

$$kurt = \frac{1}{n} \sum_{i=1}^n m_{ii}^2$$

$$m_{ij} = (X_j - \bar{X})^T S^{-1} (X_j - \bar{X})$$

$$S = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})^T$$

Obteniéndose un valor Mardia = 132.090, lo cual evidencia claramente que las variables analizadas no presentan normalidad multivariante, por consiguiente, la estimación del modelo estructural se realizó mediante la estimación de distribución libre asintótica.

Los resultados se muestran a continuación:

Diagrama de ruta del modelo de un factor (No estandarizado):

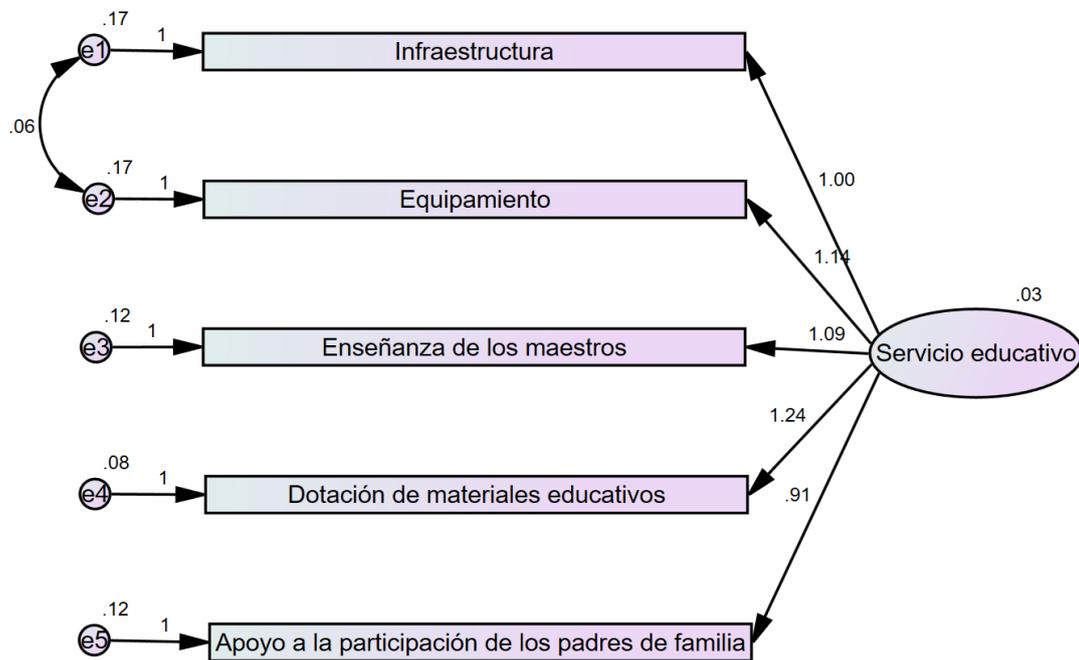
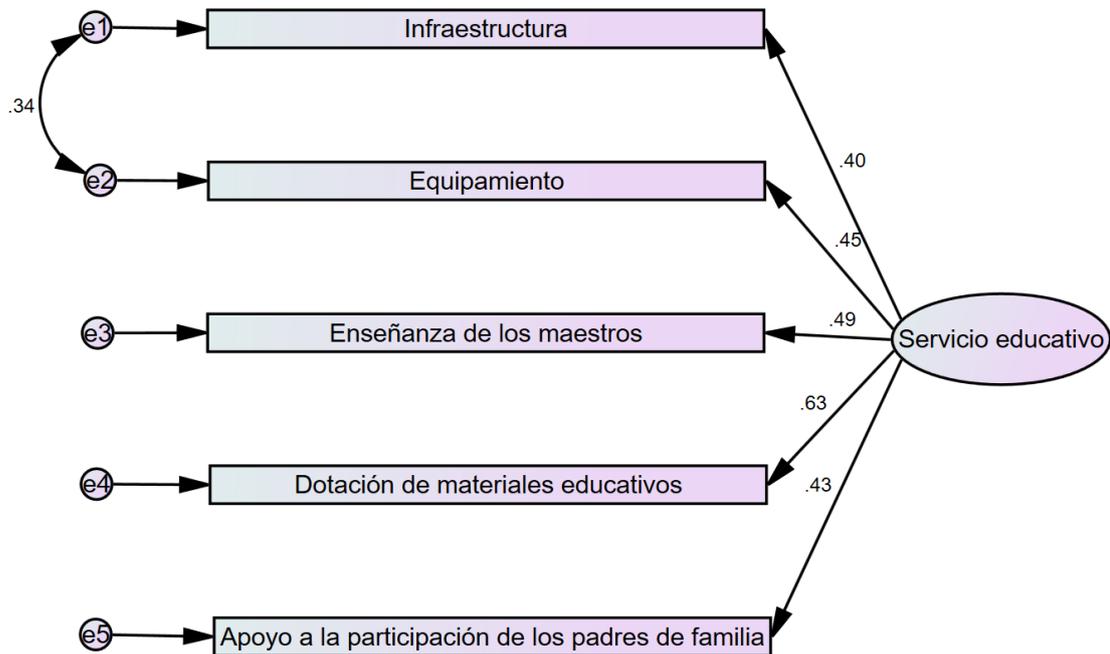


Diagrama de ruta del modelo de un factor (Estandarizado):



Así los modelos quedan descritos de la siguiente manera:

$$\text{Infraestructura} = 1.00 \times \text{ServicioEducativo} + 0.17$$

$$\text{Equipamiento} = 1.14 \times \text{ServicioEducativo} + 0.17$$

$$\text{EnseñanzaMaestros} = 1.09 \times \text{ServicioEducativo} + 0.12$$

$$\text{MaterialesEducativos} = 1.24 \times \text{ServicioEducativo} + 0.08$$

$$\text{ParticipaciónPadres} = 0.91 \times \text{ServicioEducativo} + 0.12$$

Estimación del modelo:

Número de parámetros conocidos : 15

Número de parámetros desconocidos : 11

Diferencia : 4

Por lo que se dice que el modelo se encuentra sobreidentificado.

Grados de libertad (g.l.) = 4

Índice descriptivo de bondad de ajuste:

$$\chi^2 = 6.602, \text{ sig } (p) = 0.159$$

$$H_0: S = \Sigma$$

$$H_1: S \neq \Sigma$$

Por lo que el índice descriptivo de bondad de ajuste indica que el modelo reproduce el comportamiento de los datos observados de un modo razonablemente bueno.

Índice comparativo χ^2 / gl :

Como el valor $\frac{\chi^2}{gl} = 1.65$, es menor que 2, entonces significa que el modelo presenta un buen ajuste.

Error cuadrático medio de aproximación (RMSEA):

RMSEA = 0.015, lo cual indica que el modelo estimado presenta un buen ajuste.

Residuo cuadrático medio (RMR):

RMR = 0.002, lo que indica que el modelo estimado presenta un buen ajuste.

Índice de ajuste normado (NFI):

NFI = 0.979, lo que representa un buen ajuste del modelo y representa la reducción proporcional en la función de ajuste cuando se pasa del modelo nulo al propuesto.

Índice de ajuste comparativo (CFI):

CFI = 0.991, lo que representa un buen ajuste del modelo estimado.

Índice de bondad de ajuste ajustado (AGFI):

AGFI = 0.985, lo que representa un buen ajuste del modelo estimado.

Índice de bondad de ajuste (GFI):

GFI = 0.996, lo que representa un buen ajuste del modelo estimado.

Pesos de la estimación:

	Estimate	S.E.	C.R.	P	Label
P308B1 <--- F1	1.000				
P308B2 <--- F1	1.145	.084	13.551	****	
P308B3 <--- F1	1.086	.099	10.947	***	
P308B4 <--- F1	1.242	.111	11.205	***	
P308B5 <--- F1	.910	.092	9.903	***	

Los coeficientes estimados son significativos inclusive al 0.1% de significación estadística.

Pesos de la estimación estandarizado:

	Estimate
P308B1 <--- F1	.402

	Estimate
P308B2 <--- F1	.449
P308B3 <--- F1	.494
P308B4 <--- F1	.631
P308B5 <--- F1	.433

Como se observa, la variable (P308B4) “dotación de materiales educativos” se encuentra mejor ponderado frente a las demás variables.

	Estimate	S.E.	C.R.	P	Label
e1 <--> e2	.058	.006	10.237	***	

El modelo covariado que resulta significativo es entre e1 y e2 con un coeficiente no estandarizado de 0.058, mientras que la misma relación covariada presenta un coeficiente estandarizado igual a 0.341.

	Estimate
e1 <--> e2	.341

Así mismo, las varianzas del modelo confirmatorio son significativas inclusive al 99.99% de confiabilidad.

	Estimate	S.E.	C.R.	P	Label
F1	.033	.005	6.617	***	
e1	.172	.008	21.073	***	
e2	.171	.008	21.915	***	

	Estimate	S.E.	C.R.	P	Label
e3	.121	.007	17.707	***	
e4	.077	.006	12.576	***	
e5	.118	.006	18.591	***	

Árbol de clasificación respecto a los componentes de la calidad del servicio educativo frente a la lectoescritura

Para el desarrollo del árbol de clasificación se consideraron las siguientes características:

Variables explicativas:

- Infraestructura (P308B1)
- Equipamiento (P308B2)
- Enseñanza del docente (P308B3)
- Dotación de materiales educativos (P308B4)
- Apoyo a la participación de padres de familia (P308B5)
- Edad (P208A)
- Sexo (P207)
- Región (DOMINIOg)

Variable dependiente:

- ¿Sabe leer y escribir? (P302): con respuestas Si = 1 y No = 2.

Para la elaboración de los árboles de clasificación se hizo uso del método CHAID, que a continuación se detalla:

Resumen del modelo

Especificaciones	Método de crecimiento	CHAID
	Variable dependiente	Sabe leer y escribir
	Variables independientes	Infraestructura (paredes, techos, pisos), Equipamiento (Carpetas, Laboratorios, Computadoras, etc.), Enseñanza de los maestros, Dotación de materiales educativos (Libros, Láminas, maquetas, etc.), Apoyo a la participación de los padres de familia (Promueven la participación de los padres de familia, escuela de padres), Sexo, ¿Qué edad tiene en años cumplidos?, Región
	Validación	Muestra de división
	Máxima profundidad del árbol	3
	Casos mínimos en nodo padre	100
	Casos mínimos en nodo hijo	50
Resultados	Variables independientes incluidas	¿Qué edad tiene en años cumplidos?, Región, Equipamiento (Carpetas, Laboratorios, Computadoras, etc.)
	Número de nodos	17
	Número de nodos terminales	12
	Profundidad	2

La muestra de división son 80% y 20% para el entrenamiento y la comprobación respectivamente.

Costes de clasificación errónea

Observado Pronosticado

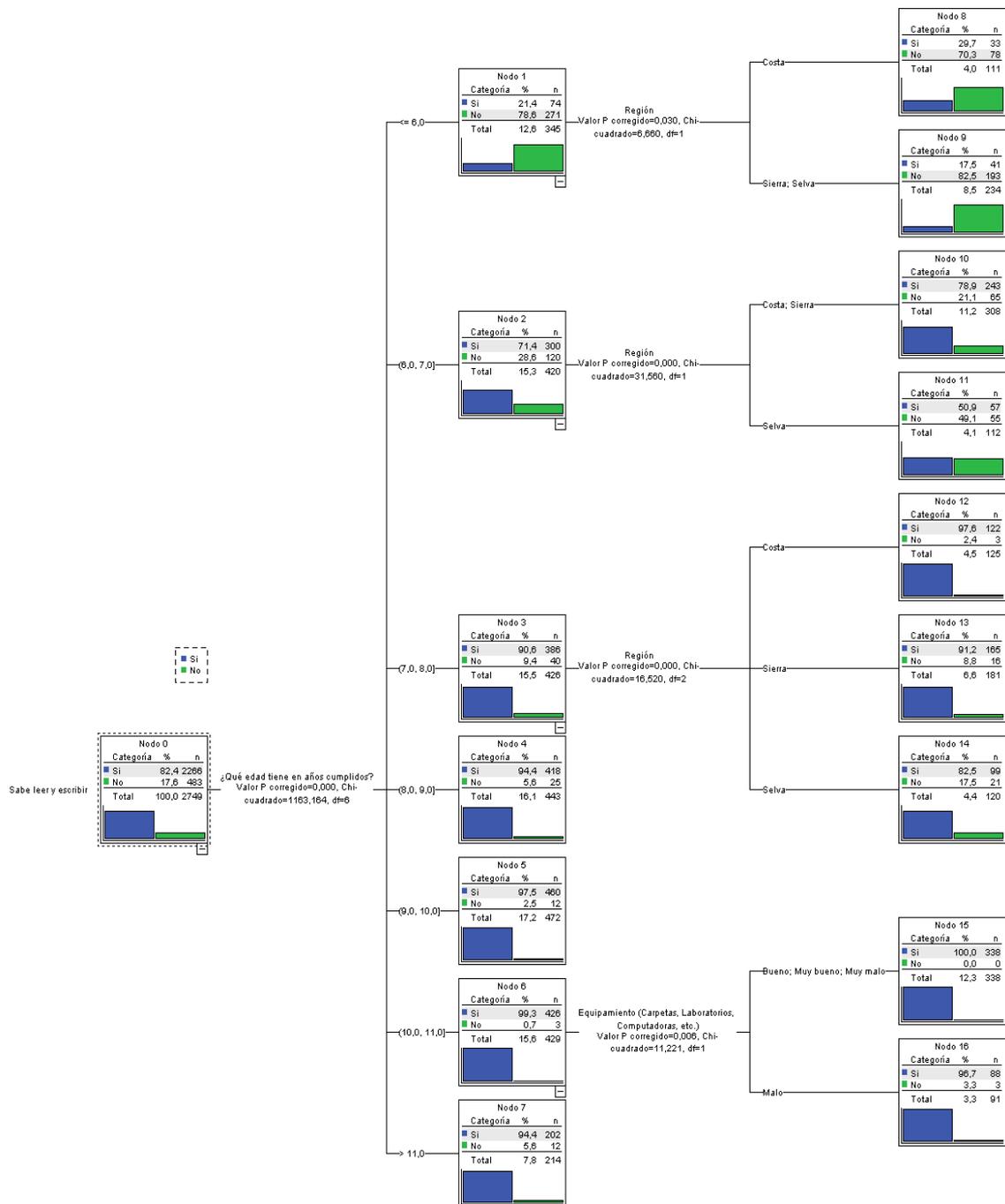
	Si	No
Si	,000	1,000
No	1,000	,000

Variable dependiente: Sabe leer y escribir

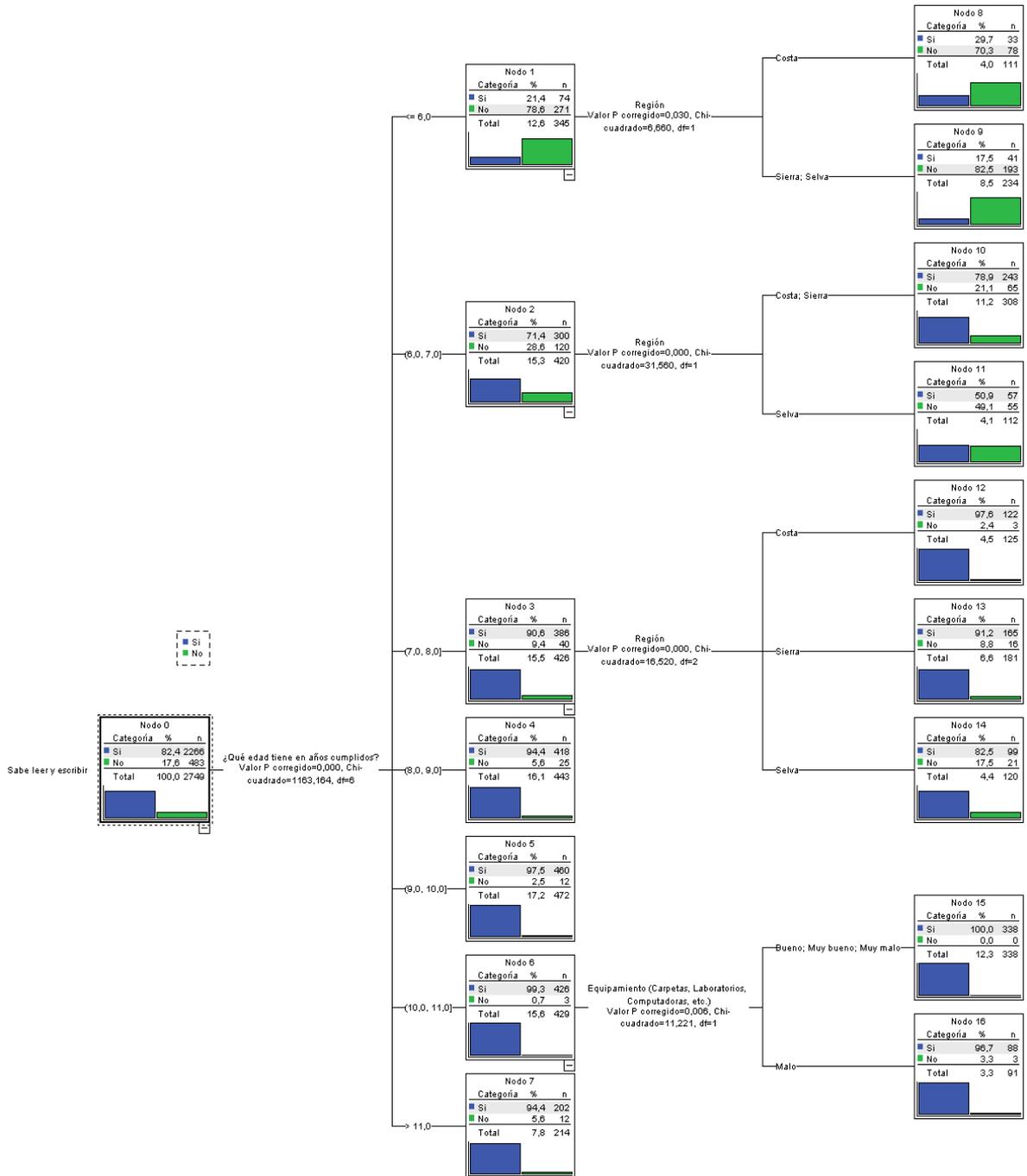
La recodificación del saber leer y escribir: Si = 0 y No = 1.

Los resultados de la muestra de entrenamiento se visualizan a continuación:

El perfil del niño que no sabe leer ni escribir esta caracterizado por tener seis o menos años de edad y pertenecer a cualquiera de las regiones (costa, sierra y selva), mientras que los niños que saben leer y escribir presentan la característica de tener más de 8 años y presentar altas calificaciones en la dimensión equipamiento del servicio educativo.



Así mismo, en la muestra de comprobación los perfiles de los niños que no saben leer ni escribir se repiten.

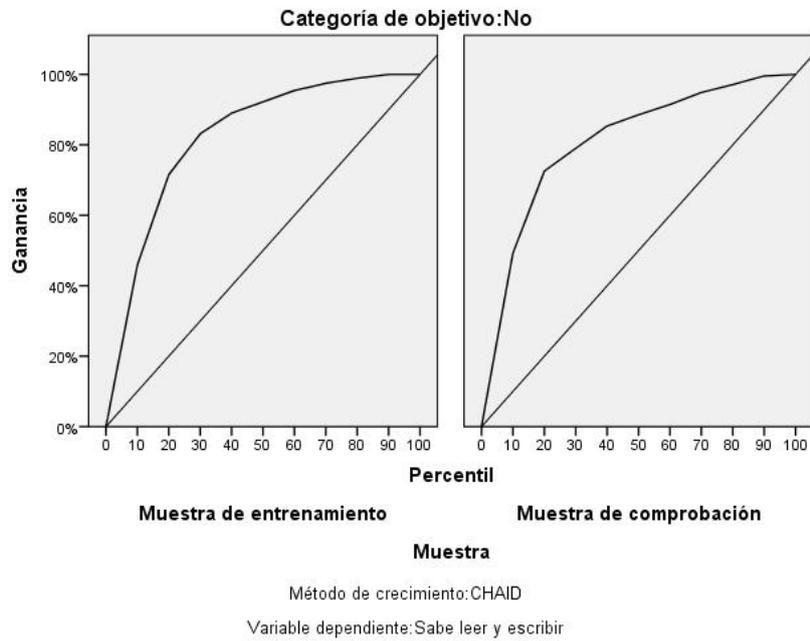


Muestra	Ganancias para nodos													
	Nodo	Nodo-por-nodo						Acumulativa						
		N	Porcentaje	N	Porcentaje	Respuesta	Índice	N	Porcentaje	N	Porcentaje	Respuesta	Índice	
Entrenamiento	9	234	8,5%	193	40,0%	82,5%	469,4%	234	8,5%	193	40,0%	82,5%	469,4%	
	8	111	4,0%	78	16,1%	70,3%	399,9%	345	12,6%	271	56,1%	78,6%	447,1%	
	11	112	4,1%	55	11,4%	49,1%	279,5%	457	16,6%	326	67,5%	71,3%	406,0%	
	10	308	11,2%	65	13,5%	21,1%	120,1%	765	27,8%	391	81,0%	51,1%	290,9%	
	14	120	4,4%	21	4,3%	17,5%	99,6%	885	32,2%	412	85,3%	46,6%	265,0%	
	13	181	6,6%	16	3,3%	8,8%	50,3%	1066	38,8%	428	88,6%	40,2%	228,5%	
	4	443	16,1%	25	5,2%	5,6%	32,1%	1509	54,9%	453	93,8%	30,0%	170,9%	
	7	214	7,8%	12	2,5%	5,6%	31,9%	1723	62,7%	465	96,3%	27,0%	153,6%	
	16	91	3,3%	3	0,6%	3,3%	18,8%	1814	66,0%	468	96,9%	25,8%	146,8%	
	5	472	17,2%	12	2,5%	2,5%	14,5%	2286	83,2%	480	99,4%	21,0%	119,5%	
	12	125	4,5%	3	0,6%	2,4%	13,7%	2411	87,7%	483	100,0%	20,0%	114,0%	
	15	338	12,3%	0	0,0%	0,0%	0,0%	2749	100,0%	483	100,0%	17,6%	100,0%	
	Prueba	9	29	8,7%	25	44,6%	86,2%	514,2%	29	8,7%	25	44,6%	86,2%	514,2%
		8	12	3,6%	7	12,5%	58,3%	347,9%	41	12,3%	32	57,1%	78,0%	465,5%
		11	20	6,0%	8	14,3%	40,0%	238,6%	61	18,3%	40	71,4%	65,6%	391,1%
10		37	11,1%	4	7,1%	10,8%	64,5%	98	29,3%	44	78,6%	44,9%	267,8%	
14		26	7,8%	3	5,4%	11,5%	68,8%	124	37,1%	47	83,9%	37,9%	226,1%	
13		12	3,6%	1	1,8%	8,3%	49,7%	136	40,7%	48	85,7%	35,3%	210,5%	
4		59	17,7%	3	5,4%	5,1%	30,3%	195	58,4%	51	91,1%	26,2%	156,0%	
7		25	7,5%	1	1,8%	4,0%	23,9%	220	65,9%	52	92,9%	23,6%	141,0%	
16		10	3,0%	1	1,8%	10,0%	59,6%	230	68,9%	53	94,6%	23,0%	137,4%	
5		54	16,2%	2	3,6%	3,7%	22,1%	284	85,0%	55	98,2%	19,4%	115,5%	
12		22	6,6%	1	1,8%	4,5%	27,1%	306	91,6%	56	100,0%	18,3%	109,2%	
15		28	8,4%	0	0,0%	0,0%	0,0%	334	100,0%	56	100,0%	16,8%	100,0%	

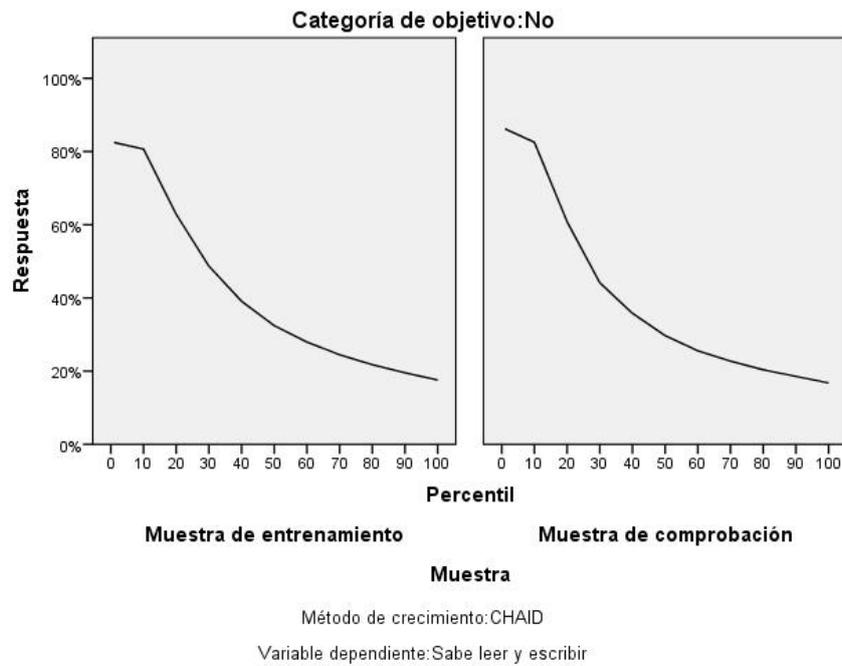
Método de crecimiento: CHAID

Variable dependiente: Sabe leer y escribir

Los índices porcentuales de ganancia reflejan que la muestra de entrenamiento y la muestra de prueba son adecuadas.



Se observa que el objetivo no sabe leer ni escribir si presenta poder de discriminación tanto para la muestra de entrenamiento y la muestra de comprobación.



Riesgo		
Muestra	Estimación	Desv. Error
Entrenamiento	,104	,006
Prueba	,099	,016

Método de crecimiento: CHAID

Variable dependiente: Sabe leer y escribir

Los niveles de riesgo son adecuados tanto para la muestra de entrenamiento como de la muestra de prueba.

Clasificación				
Muestra	Observado	Pronosticado		Porcentaje correcto
		Si	No	
Entrenamiento	Si	2192	74	96,7%
	No	212	271	56,1%
	Porcentaje global	87,4%	12,6%	89,6%
Prueba	Si	269	9	96,8%
	No	24	32	57,1%
	Porcentaje global	87,7%	12,3%	90,1%

Método de crecimiento: CHAID

Variable dependiente: Sabe leer y escribir

Finalmente, el porcentaje de clasificación correcta para la muestra de entrenamiento es de 89.6%, mientras que para la muestra de prueba el porcentaje de clasificación correcta es de 90.1%, por lo que son indicadores adecuados de la clasificación.

5.5. Evaluación

Respecto al análisis factorial confirmatorio se elaboró el modelado con un 80% (2466 registros) de los datos elegidos aleatoriamente, posteriormente se realizó el modelado del

análisis factorial confirmatorio incluyendo el 20% (617 registros) de los datos restantes, en la que se observó que en el primer modelamiento un comportamiento semejante cuando se incluyó el 100% de los datos.

De igual manera, para el análisis de los árboles de decisión se trabajó con el 80% (2466 registros) para el modelo de entrenamiento y el 20% (617 registros) para predicción, donde se predijo un 90,1% de clasificación correcta.

5.6. Implantación

Respecto al instrumento de medición de la calidad del servicio educativo peruano, cabe indicar que dicho instrumento compuesto por los 5 pilares como son la infraestructura, equipamiento, enseñanza de los maestros, dotación de los materiales educativos y el apoyo a la participación de los padres de familia forman los pilares base para tener una referencia de calidad del servicio educativo peruano, donde resalta una sub-dimensión oculta entre el equipamiento e infraestructura como activos fijos con que cuenta cada una de las instituciones, pudiendo dinamizarse los restantes tres indicadores de la calidad del servicio educativo.

Respecto al análisis de los árboles de decisión, se observa la no asociación de la capacidad de lectoescritura frente a la calificación de la calidad de servicio educativo peruano, más por el contrario, la capacidad de lectoescritura está influenciado por la edad de los niños como un factor de maduración cognitiva.

5.7. Discusión de resultados

Las habilidades de leer y escribir son necesarias para una adecuada opinión de calidad del servicio educativo que se brinda en las instituciones nacionales y privadas del país. Así, según la encuesta nacional de hogares (ENAHOG) se recoge dicha información con la finalidad de ser estudiada y analizada. Resultado de ello es la verificación de la validez de constructo por medio de la consistencia interna mediante el análisis factorial confirmatorio, mientras que los perfiles que presentan los estudiantes respecto a su habilidad de leer y escribir mediante los árboles de decisión. Esta calidad tal como lo señala Municio (2000) es una medida de ajuste del producto o servicio a los requisitos demandados por los clientes y Crosby (1980) donde señala que la calidad de un bien o servicio no es otra cosa que la conformidad del mismo con los requerimientos de los clientes, así, las necesidades de los estudiantes peruanos respecto a infraestructura, equipamiento y dotación de materiales hace que el estado peruano actúe de manera directa en la mejora y/o adecuación de las mismas, tal como lo indica García Hoz (2000) que los medios que se disponen con influyentes en la calidad de los servicios. Por otro lado, el Consejo Nacional de Educación (2007) la mejora de los aprendizajes en la educación básica depende de un conjunto de factores, que tienen que ver con el equipamiento de las instituciones, la gestión escolar, la formación de los profesores, nuevas reglas de juego para el ejercicio de la docencia, evaluación periódica de los progresos en el rendimiento de los alumnos, una educación inicial con mayor cobertura y calidad y programas que compensen las desventajas iniciales de los alumnos a nivel de su salud y nutrición. Estos últimos respecto a la enseñanza de los docentes y la participación de los padres hace que

sea activa dicha participación de los actores principales como son los docentes dentro de las aulas de enseñanza aprendizaje, así como de los padres dentro de los ambientes del hogar, dinamizando el contexto educativo de los estudiantes, por ende, la calidad de servicio educativo peruano medido por intermedio de los cinco indicadores considerados son factibles, ya que la corroboración mediante el análisis factorial confirmatorio resaltan su buen ajuste de manera lineal con el servicio educativo, con indicadores de ajuste del modelo adecuados y altos. Además, se observa que los errores aleatorios de la infraestructura se encuentran covariados con los errores aleatorios del equipamiento en los centros de enseñanza que asisten los niños.

En segundo lugar, los niños que no saben leer y escribir son caracterizados por tener seis años o menos, lo cual concuerda en que cada niño tiene su propio ritmo de aprendizaje, para poder aprender a leer y escribir deben haber conseguido antes un buen nivel de lenguaje oral y una coordinación de movimientos adecuada. Por esto, se estima que la edad ideal para el aprendizaje de la lectoescritura es en torno a los 6 años, significando que dichos niños recién inician su aprendizaje en la lectoescritura, mientras que los niños de ocho a más años ya saben leer y escribir en la mayoría de los casos de manera fluida. Sin embargo, por problemas de pobreza y dificultad de acceso a una educación adecuada se cuenta con un 17% de niños que aún no saben leer ni escribir, pese a que dichos niños debieron haber pasado por la educación inicial y al momento de la transición a la educación primaria ya se tiene noción de leer y escribir de manera silábica, sin embargo, ello no sucede en los lugares desfavorecidos de la costa, sierra y selva del país. Lo que

concuenda con lo indicado por Ferreiro en Zorzi (2003) en que la mayor parte del conocimiento que alguien puede obtener en su vida será mediante la lectura, aunque sea fuera de la escuela, entendiéndose como el paso del tiempo y el proceso propio de maduración y la necesidad de comunicación tanto verbal como escrita de las personas, tal como lo indica Braslavsky (2005) donde las actividades que se realizan en familia relacionadas con la lectura y escritura en edades tempranas asientan las bases para que los niños aprendan a leer y escribir en la escuela, donde posteriormente, será el tiempo de ocio y el placer por la lectura lo que hará que se perfeccionen y aumenten las habilidades lectoras. El ambiente donde se desenvuelve el niño debe estar relacionado con la lectura, es decir, en su entorno cuenten con la presencia de libros, revistas, documentos y el hecho que la lectura sea una actividad placentera y cotidiana en su vida

La opinión de la calidad del servicio educativo que le brinda la institución educativa al cual asiste no se encuentra relacionado ni es influenciado por la capacidad de saber leer ni escribir, ya que dicha capacidad no influye en su satisfacción por los servicios que recibe, de la cual opina libremente en la forma como lo percibe tal como lo indica Deming (2000) donde indica que la calidad es todo aquello que contribuye a la satisfacción del cliente.

**CAPÍTULO VI: CONCLUSIONES, LIMITACIONES Y
RECOMENDACIONES**

CONCLUSIONES

- La opinión sobre la calidad del servicio educativo que reciben en sus centros de enseñanza no se encuentra asociado a la capacidad que tiene los niños de leer escribir.
- Los indicadores infraestructura, equipamiento, enseñanza de los maestros, dotación de materiales educativos y apoyo a la participación de los padres de familia explican de la calidad del servicio educativo, con un buen ajuste del modelo factorial confirmatorio, además de que la infraestructura y el equipamiento se encuentran covariados.
- La capacidad de leer y escribir esta caracterizado significativamente por tener siete años de edad y asociado significativamente a pertenecer a la región costa y sierra, a los ocho años asociado a pertenecer a la costa, sierra y selva, mientras que la edad de 9, 10 y más de once asociados significativamente a saber leer y escribir. Así mismo, los niños de once años de edad saben leer y escribir y presentan opiniones de asociación significativa respecto a la calidad del servicio educativo en cuanto al equipamiento (carpetas, laboratorios, computadoras, etc), mientras que los niños que no saben leer ni escribir se encuentra caracterizado por tener 6 años de edad o menos y con asociación significativa de pertenecer a la costa, sierra o selva.

LIMITACIONES

- Limitación en cuanto a material bibliográfico, ya que en la biblioteca de la facultad de ciencias no existen libros especializados en cuanto a minería de datos.
- Debido a que la minería de datos cuenta con varias técnicas, sólo nos planteamos realizar el uso del Análisis Factorial Confirmatorio y los Árboles de Decisión, tal como están descritos en nuestros objetivos.

RECOMENDACIONES

- Al Ministerio de Educación y a las instituciones involucradas en sector educación; incidir en mejoramiento de infraestructura, equipamiento, enseñanza de los maestros, dotación de materiales educativos y el apoyo de a la participación de los padres de familia para mejorar y garantizar la calidad del servicio educativo peruano.
- A los directivos de la escuela profesional, que implementen cursos de minería de datos, para que los estudiantes tengan las herramientas adecuadas y profundicen sus conocimientos en minería de datos.

REFERENCIAS BIBLIOGRÁFICAS

Ale, J., (2005). Análisis de Clusters.

Ale, J., (2005). Introducción a Data Mining

Álvarez, A.I (2005). Hablar en español. Ed. Nobel. Oviedo

Álvarez, A.I (2005). Leer en español. Ed. Nobel. Oviedo

BERRY, T. (2002). Cómo gerenciar la transformación hacia la Calidad Total. Bogotá: McGraw-Hill

BOCYL. (3 marzo 2011). ORDEN EDU/152/2011, planes de fomento.

Bollen, K. A. (1989). Structural equations with latent variables. New York: John Wiley y Sons.

Braslavsky, B. (1991) La lengua escrita y los procesos de adquisición del conocimiento en una concepción socio histórico-cultural. Recuperado de <http://www.iacd.oas.org.com>

CANTON MAYO, I. (2000). Evaluación, cambio y calidad en los centros educativos. Buenos Aires. FUNDEC.

CARVAJAR e RAMOS (2001). Ensinar ou aprender a ler e escrever Organizado por Francisco Carvajal Pérez e Joaquim Ramos García: trad. Cláudia Schiling – Porto Alegre: Artmed Editora

Catillo, Medina y Vega (2016). Esquema de Visualización para Modelos de Clústeres en Minería de Datos. Revista Ibérica de Sistemas y Tecnologías de Información. RISTI, N.º 21, 03/2017. DOI: 10.17013/risti.21.67–84

COLOMER, T. e CAMPS, A. - (2002): Trad. Fátima Murad - Ensinar a ler, ensinar a compreender. Porto Alegre. Artimed

- COMUNIDAD ECONÓMICA EUROPEA (1995). Libro Blanco sobre educación y formación en la Unión Europea. Luxemburgo: CEE
- Correa Medina, E. J. (2007). Conciencia fonológica y percepción visual en la lectura inicial de niños del primer grado de primaria. Lima: Pontificia Universidad Católica del Perú. Obtenido de http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/421/CORREA_MEDINA_ELIZABETH_CONCIENCIA_FONOLOGICA.pdf?sequence=1&isAllowed=y
- Conejo y Carmiol (2017). Conocimientos sobre la lectoescritura emergente y prácticas en las aulas para su promoción: Un estudio con docentes de Educación preescolar en Costa Rica. *Revista Costarricense de Psicología*. Jul-Dic 2017, Vol. 36, N.º 2, p. 105-121.
- Chen, H., W. Chung, J. Xu, G. Wang, Y. Qin, M. Chau, 2004. Crime Data Mining: A General Framework and Some Examples. *IEEE Computer Society*, vol. 37, no. 4. Páginas 50-56.
- Chen, M., J. Han, 1996. Data mining: An overview from database perspective. *IEEE Transactions on Knowledge and Data Eng*
- Cheeseman, P., J. Stutz, 1996. Bayesian classification (AutoClass): Theory and results. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*. Páginas 153-180. AAAI/MIT Press, California
- CRISP-DM (2012). Manual CRISP-DM de IBM SPSS Modeler.
- Cuetos, Molina, Suárez-Coalla y Llenderozas (2017). Validación del test para la detección temprana de las dificultades en el aprendizaje de la lectura y escritura. *Rev Pediatr Aten Primaria*. 2017;19:241-6.
- CROSBY, P. (1980). *Quality is free*. New York: Penguin Books.

Cuadras Carles (2014). Nuevos Métodos de Análisis Multivariante

Edwards Deming, (2000). Out of the Crisis 2nd ed. Cambridge, Mass.: Mit Press.

Díaz, H. (7 de Marzo de 2014). El Comercio. Obtenido de Los tres problemas que existen en el sector educación del Perú: <http://elcomercio.pe/economia/peru/tres-problemas-que-existen-sector-educacion-peru-noticia-1714189>

Dubois, M. E. (2000). El proceso de la lectura: de la teoría a la práctica. . Buenos Aires: Aique

Elder IV, J., D. Pregibon, 1996. A statistical perspective on knowledge discovery in databases. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, Advances in Knowledge Discovery and Data Mining. Páginas 83-115. AAAI/MIT Press, California, EEUU.

Estrada, Zamarripa, Zúñiga y Martínez (2016). Aportaciones desde la minería de datos al proceso de captación de matrícula en instituciones de educación superior particulares. Revista Electrónica Educare (Educare Electronic Journal) Vol. 20(3) SETIEMBRE-DICIEMBRE, 2016: 1-21. DOI: <http://dx.doi.org/10.15359/ree.20-3.11>

Evangelos, S., J. Han, 1996. Proceeding of the Second International Conference on Knowledge Discovery and Data Mining. Portland, EEUU.

Essenreiter R., M. Karrenbach, S. Treitel, 1999. Identification and classification of multiple reflections with selforganizing maps. Geophysical Institute, University of Karlsruhe, Germany

Ferreiro, E. (3 de febrero de 2011). Las ferrerianas. Obtenido de <http://lasferrerianas5.blogspot.com.es/>

FERREIRO, E. (1982) - en - ZORZI, J. L. (2003):Temas em Educação- Livro das jornadas 2003 – Edit. Futuro Congresso e Eventos- São Paulo

- Fisher, D., 1996. Iterative optimization and simplification of hierarchical clusterings. Department of Computer Science. Vanderbilt University, Nashville, EEUU
- GARCÍA HOZ, V. (2000). Calidad de la Educación, trabajo y libertad. Madrid: Editorial Dossat S.A
- Gelp, I. (1987). Historia de la escritura. Alianza, Madrid, España.
- Hair, Joseph, Anderson, Rolph, Tatham, Ronald & Black, William (1999). Análisis multivariante. Prentice Hall. Madrid, España.
- Han, J., M. Kamber, 2001. Data mining: Concepts and techniques. Morgan Kauffmann Publishers
- Hand, D., H. Mannila, P. Smyth, 2001. Principles of data mining. The MIT Press, California, EEUU.
- Hernández Orallo, J., 2000. Extracción Automática de Conocimiento de base de datos e ingeniería del software. Programación declarativa e ingeniería de la programación
- Hincapié (2017). Didáctica de la lectoescritura universitaria: dos reseñas a Jonathan Potter. Praxis & Saber - Vol. 8. Núm. 18 - Setiembre - Diciembre 2017- Pág. 133 - 156. DOI: <https://doi.org/10.19053/22160159.v8.n18.2017.4318>
- ITALO Galvino, en VARGAS, S. (1995): Leitura: Uma Aprendizagem de Prazer, 3ª edição – Rio de Janeiro: José Olympio
- Jain, A. K., R. C. Dubes, 1988. Algorithms for Clustering Data. Prentice Hall
- Joseph M. Juran, (1967). Management of Quality Control. New York: Mc Graw-Hill
- Kaoru Ishikawa, (1986). Qué es el Control total de calidad? La modalidad Japonesa. Bogotá: Editorial Norma.
- Kaufman, L., P. J. Rousseeuw, 1990. Finding Groups in Data: an Introduction to Cluster Analysis. Wiley-Interscience

- Kohonen, T., 1995. *Self-Organizing Maps*. Springer-Verlag
- Lior Rokach, Oded Maimon (2008). *Data mining with decision trees: theory and applications*.
- Lu H., R. Setiono, H. Liu, 1995. *NeuroRule: A Connectionist Approach to Data Mining*. Proceedings of the 21th International Conference on Very Large Data Bases. Páginas 478/489.
- Lucas, V. (2014). *La Lectoescritura en la Etapa de Educación Primaria*. Universidad de Valladolid. Escuela de Educación de Soria
- Lu H., R. Setiono, H. Liu, (1995). *NeuroRule: A Connectionist Approach to Data Mining*. Proceedings of the 21th International Conference on Very Large Data Bases. Páginas 478/489
- Magán-Hervás y Gértrudix-Barrio (2017). Influencia de las actividades audio-musicales en la adquisición de la lectoescritura en niños y niñas de cinco años. *Revista Electrónica Educare (Educare Electronic Journal)*. Vol. 21(1) ENERO-ABRIL, 2017: 1-22. doi: <http://dx.doi.org/10.15359/ree.21-1.15>
- Mannila, H., (1997). *Methods and problems in data mining*. In Proc. of International Conference on Database Theory, Delphi, Greece
- Martí, E. (2003). *Representar el mundo externamente. La adquisición infantil de los sistemas externos de representación*. Madrid: Visor
- MARTINS, M. H. (2004): *O que é leitura* São Paulo: Brasiliense- (Coleção Primeiros Passos; 74)
- Mendoza y Rojas (2016). Resultados concluyentes del estudio sobre el nivel de lectoescritura en cadetes de la Escuela Militar (2010-2014). *Revista Científica "General José María Córdova"*. Volumen 14, Número 18 (julio-diciembre) Año 2016.

- Michalski R., 1983. A Theory and Methodology of Inductive Learning. MorganKauffman, EEUU
- Montealegre, R., & Forero, L. A. (2006). Desarrollo de la lectoescritura: adquisición y dominio. *Acta Colombiana de Psicología*, 9(1), 25-40.
- Morales, E., 2003. Descubrimiento de Conocimiento en Bases de Datos
- MUNICIO, P. (2000). Herramientas para la Evaluación de la Calidad. *Mecanografías Escuela Española, CISS Praxis: Madrid*
- Ng, R., J. Han, 1994. Efficient and effective clustering method for spatial data mining. In *Proc. 1994 Int. Conf. Very Large Data Bases*, (Páginas 144/155), Santiago de Chile, Chile
- Ochoa, Orellana, Sánchez y Dávila (2014). Componente web para el análisis de información clínica usando la técnica de Minería de Datos por agrupamiento. *Revista Cubana de Informática Médica* 2014:6(1)5-16. <http://scielo.sld.cu>
- ORLANDI, E.P. (2001): *Discurso e Leitura / – 6ª edição – São Paulo, Cortez; Campinas, S P: Editora da Universidade Estadual de Campinas – Coleção passando a limpo.*
- Palos, Ávalos, Flores y Montes (2017). Creencias de madres y docentes sobre el aprendizaje de la lectoescritura en Educación Preescolar. Creencias de madres y docentes sobre el aprendizaje de la lectoescritura en Educación Preescolar. *Revista Electrónica “Actualidades Investigativas en Educación”*. Volumen 17 Número 3, Año 2017. DOI: <http://dx.doi.org/10.15517/aie.v17i3.29066>
- Pedro Isasi Viñuela; Inés M. Galvan León (2004). *Las redes neuronales artificiales: un enfoque práctico.* Prentice Hall.
- Pérez López, César & Santín González, Daniel (2007). *Minería de datos.* Thomson Ediciones Paraninfo S.A. Madrid, España.

- Perichinsky, G. y R. Garcia Martínez, 2000. A Data Mining Approach to Computational Taxonomy. Proceedings del Workshop de Investigadores en Ciencias de la Computación. Páginas 107-110. Editado por Departamento de Publicaciones de la Facultad de Informática. Universidad Nacional de La Plata, Buenos Aires, Argentina
- Perichinsky, G., R. García-Martínez, A. Proto, 2000. Knowledge Discovery Based on Computational Taxonomy And Intelligent Data Mining. CD del VI Congreso Argentino de Ciencias de la Computación. Ushuaia, Argentina
- Perichinsky, G., R. Garcia-Martínez, A. Proto, A. Sevetto, D. Grossi, 2001. Integrated Environment of Systems Automated Engineering. Proceedings del II Workshop de Investigadores en Ciencias de la Computación. Editado por Universidad Nacional de San Luis en el CD Wicc 2001.
- Perichinsky, G., M. Servente, A. Servetto, R. García-Martínez, R. Orellana, A. Plastino, 2003. Taxonomic Evidence and Robustness of the Classification Applying Intelligent Data Mining. Proceedings del VIII Congreso Argentino de Ciencias de la Computación. Pág. 1797-1808.
- Perversi, I. (2007). Aplicación de minería de datos para la exploración y detección de patrones delictivos en Argentina. Argentina: Instituto Tecnológico de Buenos Aires.
- Piatetski-Shapiro, G., U. Fayyad, P. Smith, 1996. From data mining to Knowledge discovery. AAAI Press/MIT Press, California, EEUU.
- Piatetski-Shapiro, G.; W. Frawley, C. Matheus, 1991. Knowledge discovery in databases: an overview. AAAI-MIP Press, Menlo Park, California, EEUU.
- Quinlan, J., 1993c. C4.5: Programs for Machine Learning. Morgan Kaufmann
- Rigal, R. (2006). Educación motriz y educación psicomotriz en Preescolar y Primaria. Inde publicaciones.

- Ruiz Torres, Mary Karina (2007). Data warehouse y minería de datos. Ed. Dirección General de Servicios de Cómputo Académico, México.
- Sastre-Gómez, Celis-Leal, Roa-De La Torre y Luengas-Monroy (2017). La conciencia fonológica en contextos educativos y terapéuticos: efectos sobre el aprendizaje de la lectura. *Educación y Educadores*, 20(2), 175-190. DOI: 10.5294/edu.2017.20.2.1
- Servente, M., R. García-Martínez, 2002. Tesis Doctoral Algoritmos TDIDT aplicados a la minería de datos inteligente. Universidad de Buenos Aires, Argentina
- SILVA, E. T. da (1997). *Leitura & realidade brasileira*. (5ª Ed) Porto Alegre: Mercado Aberto 160p
- SOLE, I. (1998). *Estratégias de Leitura*. Trad: Cláudia Schilling (6ª ed). Porto Alegre: Artmed.
- Soto (2017). El cuento como mediación pedagógica para el fortalecimiento de la lectoescritura. *Revista del Instituto de Estudios en Educación Universidad del Norte*. N° 27 julio-diciembre, 2017. DOI: <http://dx.doi.org/10.14482/zp.27.10979>
- Valerio Jiménez, M. (15 de Febrero de 2011). Problemas de lectoescritura. *Revista de Claseshistoria*, 2-11. Obtenido de <https://dialnet.unirioja.es/descarga/articulo/5170654.pdf>
- Vera, J. (2014). Los servicios educativos. *Revista de la Asociación de Inspectores de Educación de España*. 2014, N° 22. pp 1-13.
- Vesanto J., E. Alhoniemi, 2000. Clustering of the Self-Organizing Map. *IEEE transactions on neural networks*, Vol 11, No. 3.
- Visbal, Mendoza y Orjuela (2017). Predicción de la eficiencia de las instituciones de educación superior colombianas con análisis envolvente de datos y minería de

datos. pensamiento & gestión, 42. Universidad del Norte, 40-161, 2017. DOI:
<http://dx.doi.org/10.14482/pege.41.9704>

VILLARROEL, C. (1997). Material de evaluación y acreditación de la E. S. venezolana.

Vosniadou, S. (2006). *Cómo aprenden los niños*. Serie Prácticas Educativas n° 7. México:
Universidad Pedagógica Nacional.

Zeithaml, V. & Parasuraman, A. (2004). Relevant knowledge series: Service Quality.
Cambridge, Mass.: Marketing Science Institute.

Zhang, T., R. Ramakrishnan, M. Livny. (1996). BIRCH: an efficient data clustering
method for very large databases. In Proc. 1996 ACM-SIGMOD Int. Conf.
Management of Data, Montreal, Canadá

ANEXOS

311. EN LOS ÚLTIMOS 12 MESES, DE..... A....., ¿GASTÓ, OBTUVO, CONSIGUIÓ O LE REGALARON:

¿Cuántas veces realizó el gasto durante los últimos 12 meses?

<p>A. ¿Cómo obtuvo el(la)</p> <p>Comprado?.....1 Autoconsumo?.....2 Autosuministro?.....3 Pago en especie?.....4 Otros hogares?.....5 Programa Social.....6 Otro? (especifique).....7</p> <p style="text-align: center;">PASE A "D"</p>	<p>B.</p> <p>¿Cuánto fue el monto de</p>	<p>C.</p> <p>¿Dónde compró el(los):</p> <p style="text-align: center;">PASE A "E"</p>	<p>D.</p> <p>¿Cuánto cree que le costaría si tuviera que pagar por este producto o servicio?</p> <div style="border: 1px solid black; padding: 5px; text-align: center;"> SI EN 311-A CIRCULÓ EL CÓDIGO 2, 3, 4, 5, 6 y/o 7 </div>	<p>E.</p> <p>Número de veces</p>
--	---	--	---	---

	Sí	No	Código							S/.	Código	MONTO S/.						
												311-A=2	311-A=3	311-A=4	311-A=5	311-A=6	311-A=7	
												Enteros	Enteros	Enteros	Enteros	Enteros	Enteros	
1. Uniforme escolar?	1	2	1	2	3	4	5	6	7									
2. Calzado escolar?	1	2	1	2	3	4	5	6	7									
3. Libros y textos?	1	2	1	2	3	4	5	6	7									
4. Útiles escolares?	1	2	1	2	3	4	5	6	7									
5. Matrícula?	1	2	1	2	3	4	5	6	7									
6. APAFA?	1	2	1	2	3	4	5	6	7									
7. Otros (fotocopias, cuotas extraordinarias, etc.? _____ (Especifique)	1	2	1	2	3	4	5	6	7									
8. Laptop del Programa "Una Laptop por Niño"	1	2						6										
TOTAL																		

1. Ambulante
2. Librería
3. Centro de estudios
4. Feria
5. Puesto de mercado
6. Bazar
7. Bodega, tienda, etc.
8. Tienda especializada al por mayor
9. Tienda especializada al por menor
10. Otro (especifique)

4

OBSERVACIONES

312. EN EL MES ANTERIOR, GASTÓ EN:

¿Cuántas veces realizó el gasto durante los últimos 12 meses?

		A. ¿Cómo obtuvieron el(la):						B. ¿Cuánto fue el monto de	C. ¿Cuánto cree que le costaría si tuviera que pagarlo?					D. Número de veces		
		Comprado?..... 1 Autosuministro?2 Pago en especie?3 Otros hogares?. 4 Programa social?5 Otro(especifique)6					 ?	MONTO S/.							
		Código						S/.	312-A=2	312-A=3	312-A=4	312-A=5	312-A=6			
									Enteros	Enteros	Enteros	Enteros	Enteros			
1.	Pensión mensual de enseñanza?	1	2	1	2	3	4	5	6.....							
2.	Movilidad particular al centro de enseñanza?	1	2	1	2	3	4	5	6.....							
TOTAL																

1. Pensión mensual de enseñanza?

2. Movilidad particular al centro de enseñanza?

TOTAL

B. ENCUESTADOR: Transcriba la edad de pregunta 208, capítulo 200.

Menores de 25 años: Transcriba el código de pgta. 306 y 307

Si alguna de ellas es igual a 2 continúe con 313A

Si ambas son iguales a 1

De 25 años a más → **PASE A 314A**

313A. ¿CUÁL ES LA PRINCIPAL RAZÓN POR LA QUE NO ESTÁ MATRICULADO O NO ASISTE A ALGÚN CENTRO O PROGRAMA DE EDUCACIÓN BÁSICA O SUPERIOR?

- | | |
|---|--|
| Problemas económicos..... 1 | No existe centro de educación básica o superior en el centro poblado 7 |
| Estoy trabajando 2 | Asiste a un centro de Educación Técnico Productiva..... 8 |
| Terminó sus estudios: secundarios/ superiores /asiste a academia preuniversitaria 3 | No me interesa/no me gusta el estudio 9 |
| No tiene la edad suficiente (para el grupo 3 -5 años)..... 4 | Se dedica a los quehaceres del hogar..... 10 |
| Problemas familiares..... 5 | Otra razón 11 |
| De vacaciones 6 | (Especifique) |

SERVICIO DE INTERNET (Para las personas de 6 años y más de edad)

314A. EN EL MES ANTERIOR, ¿UD. HIZO USO DEL SERVICIO DE INTERNET?

Sí..... 1

No 2 → **PASE A CAP. 400**

<p>314B. EN EL MES ANTERIOR, ¿USÓ UD. EL SERVICIO DE INTERNET EN:</p> <p>(Acepte una o más alternativas)</p> <p>El hogar? 1</p> <p>El trabajo? 2</p> <p>Un establecimiento educativo? 3</p> <p>Una cabina pública? 4</p> <p>En casa de otra persona? 5</p> <p>Otro? 6</p> <p>(Especifique)</p> <p>Acceso móvil a internet? 7</p>	<p>314B1. EN EL MES ANTERIOR, ¿EL SERVICIO DE INTERNET LO USÓ A TRAVÉS DE UNA/UN:</p> <p>(Acepte una o más alternativas)</p> <p>Computadora? 1</p> <p>Laptop? 2</p> <p>Celular propio (con o sin plan de datos)? 3</p> <p>Celular de un familiar o amigo? 4</p> <p>Celular de su centro de trabajo? 5</p> <p>Tablet ? 6</p> <p>Otro? 7</p> <p>(Especifique)</p>	<p>314C. EN EL MES ANTERIOR, ¿DÓNDE LO USÓ CON MAYOR FRECUENCIA?</p> <p><input type="text"/></p> <p>Código</p> <p>Anote el código de P314B)</p>	<p>314D. ¿UD. USA INTERNET AL MENOS:</p> <p>(Referido al uso de mayor frecuencia)</p> <p>Una vez al día? 1</p> <p>Una vez a la semana? 2</p> <p>Una vez al mes? 3</p> <p>Cada 2 meses o más? 4</p>
---	--	---	---

ENCUESTADOR: Si en la pregunta 314B alguna de las alternativas que circuló corresponde al código 4 "Cabina Pública", continúe con la pregunta 315. De lo contrario pase a la pregunta 316.

315. En el mes anterior ¿Obtuvo el servicio de Internet en cabina pública:		315A. ¿Cuánto fue el monto del servicio en el mes anterior?	315B. ¿Cuánto cree que le costaría si tuviera que pagar por este servicio que usó el mes anterior?							
Comprado?..... 1		S/.	MONTO S/.							
Autosuministro? 2			315=2	315=3	315=4	315=5	315=6			
Pago en especie? 3			Enteros	Enteros	Enteros	Enteros	Enteros			
Otros hogares? 4										
Programa social? 5										
Otro? (especifique)..... 6										
Código										
1	2	3	4	5	6.....					

OBSERVACIONES

1

316. ¿USÓ EL INTERNET PARA: (Acepte una o más alternativas)		
	Si	No
1. Obtener información (sobre bienes y servicios, salud, organizaciones gubernamentales)?	1	2
2. Comunicarse (e-mail, chat, llamadas por Skype, WhatsApp, Facebook, Twitter, etc.)?	1	2
3. Comprar productos y/o servicios?	1	2
4. Operaciones de banca electrónica y otros servicios financieros?	1	2
5. Educación formal y actividades de capacitación?	1	2
6. Transacciones (interactuar) con organizaciones estatales/autoridades públicas?	1	2
7. Actividades de entretenimiento (juegos de video, obtener películas, música, videos, escuchar radio, leer periódico, etc.)?	1	2
8. Vender productos y/o servicios (Mercado Libre, OLX, Facebook, etc.)?	1	2
9. Otros?(Especifique).....	1	2
10. Otros?(Especifique).....	1	2
11. Otros?(Especifique).....	1	2

MODULO 300A: RESULTADO DE LA EVALUACIÓN CENSAL DE ESTUDIANTES 2015 Y SATISFACCIÓN DE LOS PADRES DE FAMILIA CON LA INSTITUCIÓN EDUCATIVA DONDE ESTUDIAN SUS HIJOS.
(Sólo para el Padre, Madre o Apoderado de los niños que el año 2015, cursaron el 2do grado de educación primaria)

Persona N°	Nombre :	Informante N°
------------	----------	---------------

317. EN LA INSTITUCIÓN EDUCATIVA DONDE
 ...^(Nombre del niño) **ASISTIÓ AL 2DO GRADO EL AÑO 2015 SE LE ENSEÑÓ A LEER Y ESCRIBIR EN SU LENGUA MATERNA?**

Si..... 1
 No..... 2

“Sr. (Sra.), el año pasado el ministerio de educación realizó la evaluación censal para los alumnos del 2º grado de primaria, acerca de lógico matemática y comprensión de lectura (Mostrar los exámenes)”.

Encuestador: No olvide mostrar al informante el “Reporte Individual de Resultados”.

319. ¿HA RECIBIDO UD. EL REPORTE INDIVIDUAL CON LOS RESULTADOS QUE OBTUVO
 ...^(Nombre del niño) **EN LA EVALUACIÓN CENSAL DE ESTUDIANTES REALIZADA EN NOVIEMBRE DEL 2015?**

Si..... 1 → ¿En qué mes recibió los resultados?
 Mes

No 2 → **PASE AL CAP. 400**

318. ¿...^(Nombre del niño) **HA PARTICIPADO EN LA EVALUACIÓN CENSAL DE ESTUDIANTES REALIZADA POR EL MINISTERIO DE EDUCACIÓN EN NOVIEMBRE DEL 2015?**

Si..... 1
 No..... 2 → **PASE AL CAPÍTULO 400.**

320. ¿CÓMO LE ENTREGARON EL REPORTE INDIVIDUAL DE LA EVALUACIÓN CENSAL DE ESTUDIANTES 2015 DE :
 (Nombre del niño)

El niño(a) lo trajo de la escuela? 1
Se lo entregó a Ud. de manera personal el profesor o director de la escuela? 2
Se lo entregaron a Ud. en una reunión de padres de familia, pero no le explicaron el contenido del reporte? 3
Se lo entregaron a Ud. en una reunión de padres de familia especialmente organizada para explicar el contenido del reporte? 4
Otra forma? 5
 (Especifique)

OBSERVACIONES	