

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

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



**Algoritmo CHAID aplicado a los predictores de las víctimas de robo  
del Perú, según la Encuesta Nacional Especializada sobre  
Victimización 2017**

**Tesis para Optar el Título de:**

**Licenciado en Estadística e Informática**

**Bach. DARWIN JAMES ZAVALA SALVADOR**

**Bach. WANG MARCO ALVARADO LÁZARO**

**Asesor: Dr. Ángel Deciderio Mendoza López**

**HUARAZ - ANCASH - PERÚ**

**2019**

## **DEDICATORIA**

A Dios, a nuestros padres, hermanos y familiares quienes con su apoyo y motivación incondicional impulsaron la culminación de nuestra carrera profesional.

Darwin y Wang

## **AGRADECIMIENTO**

El agradecimiento a la Universidad Nacional Santiago Antúnez de Mayolo, docentes y autoridades de la Escuela Académico Profesional de Estadística e Informática, que nos supieron orientar en nuestra vida académica y profesional para ser hombres de bien. Así mismo, el agradecimiento infinito a todos y cada uno de nuestros amigos y compañeros universitarios, quienes siempre nos brindaron su amistad y apoyo permanente y constante para realizar y culminar nuestra tesis.

Darwin y Wang

## HOJA DE VISTO BUENO

---

MSc. Jorge Luis Llanos Tiznado

Presidente

---

Dr. Roger Pedro Norabuena Figueroa

Secretario

---

Lic. Hugo Walter Maldonado Leyva

Vocal

## INDICE

DEDICATORIA .....	I
AGRADECIMIENTO .....	II
RESUMEN .....	VI
ABSTRACT .....	VII
INTRODUCCION .....	VIII
<b>CAPITULO I: EL PROBLEMA .....</b>	<b>10</b>
1.1. SELECCIÓN DEL PROBLEMA .....	11
1.2. JUSTIFICACION DEL ESTUDIO .....	13
1.3. OBJETIVOS DE LA INVESTIGACION .....	14
1.3.1. <i>OBJETIVO GENERAL</i> .....	14
1.3.2. <i>OBJETIVOS ESPECIFICOS</i> .....	14
<b>CAPITULO II: MARCO TEORICO CONCEPTUAL .....</b>	<b>15</b>
2.1. ANTECEDENTES DEL PROBLEMA .....	16
2.2. PLANTEAMIENTO TEORICO ATINGENTE .....	24
2.2.1. <i>BASES TEORICAS</i> .....	24
A. ALGORITMO CHAID .....	24
B. PREDICTORES DE LAS VICTIMAS DE ROBO .....	39
2.2.2. <i>DEFINICION DE TERMINOS</i> .....	42
<b>CAPITULO III: HIPOTESIS Y OPERACIONALIZACION DE VARIABLES .....</b>	<b>43</b>
3.1. HIPOTESIS .....	44
3.2. OPERACIONALIZACION DE VARIABLES .....	44
<b>CAPITULO IV: METODOLOGIA .....</b>	<b>45</b>
4.1. MATERIALES Y LUGAR DE EJECUCION .....	46
4.1.1. <i>LUGAR</i> .....	46
4.1.2. <i>MATERIALES</i> .....	46
4.1.3. <i>EQUIPOS</i> .....	46
4.1.4. <i>RECURSO HUMANO</i> .....	46
4.2. METODOS .....	46
4.2.1. <i>TECNICAS DE RECOLECCION DE DATOS</i> .....	46
4.2.2. <i>TIPO DE ESTUDIO</i> .....	46
4.2.3. <i>DISEÑO DE LA INVESTIGACION</i> .....	47
4.3. POBLACION Y MUESTRA .....	47
4.3.1. <i>POBLACION</i> .....	47
4.3.2. <i>MUESTRA</i> .....	47
4.4. INSTRUMENTOS DE RECOPIACION DE DATOS .....	47
4.5. ANALISIS ESTADISTICO E INTERPRETACION DE LA INFORMACION .....	48
<b>CAPITULO V: ANALISIS Y DISCUSION DE RESULTADOS .....</b>	<b>49</b>
5.1. ANALISIS DE LOS DATOS .....	50
<i>EVALUACION</i> .....	79
<i>IMPLANTACION</i> .....	80
5.2. DISCUSION DE RESULTADOS .....	80

<b>CAPITULO VI: CONCLUSIONES, LIMITACIONES Y RECOMENDACIONES-----</b>	<b>83</b>
<b>CONCLUSIONES-----</b>	<b>84</b>
<b>LIMITACIONES-----</b>	<b>85</b>
<b>RECOMENDACIONES-----</b>	<b>86</b>
<b>REFERENCIAS BIBLIOGRAFICAS-----</b>	<b>87</b>
<b>ANEXOS-----</b>	<b>91</b>

## RESUMEN

El presente trabajo de tesis según su diseño es no experimental, transversal de tipo correlacional y tuvo por objetivo conocer los predictores de las víctimas de robo del Perú mediante el algoritmo CHAID, según la Encuesta Nacional Especializada sobre Victimización 2017. Para ello se contó con una muestra de 37, 820 personas de 15 años de edad a más años.

La aplicación del Algoritmo CHAID a los predictores de las víctimas de robo del Perú, según la Encuesta Nacional Especializada sobre Victimización 2017 dio como resultado el perfil de la víctima de robo en Perú más riesgosa y está caracterizado por: presentar una edad alta de riesgo de robo (menor a 18 años de edad y mayores a 40 años de edad), sentirse inseguro al caminar por la calle, que si han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado y mala calificación del trabajo de la comisaria.

Palabras Clave: Algoritmo CHAID, minería de datos, robo.

## **ABSTRACT**

The present thesis work according to its design is non-experimental, cross-correlated type and aimed to know the predictors of victims of theft of Peru through the CHAID algorithm, according to the National Specialized Survey on Victimization 2017. For this purpose, there was a sample of 37, 820 surveys.

The application of the CHAID Algorithm to the predictors of victims of robbery in Peru, according to the National Specialized Survey on Victimization 2017, resulted in the profile of the riskiest robbery victim in Peru and is characterized by: presenting a high risk age of robbery (under 18 years of age and over 40 years of age), feeling insecure when walking down the street, if they have stopped carrying a lot of cash, with higher education level than postgraduate and bad qualification of the work of the commissary.

**Keywords:** CHAID Algorithm, data mining, theft.



## INTRODUCCION

El presente trabajo de tesis es el resultado de mucha dedicación, esfuerzo y tiempo en la búsqueda de los datos, enlace de bases de datos, la limpieza de las tablas y bases de datos y posterior procesamiento de los datos, para ser analizados mediante el algoritmo CHAID y 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 del uso del algoritmo CHAID. 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 estudiada.

En el capítulo 4, se detallan aspectos metodológicos del presente trabajo de investigación. Así, el lugar de ejecución fue Instituto Nacional de Estadística e Informática, con diseño de investigación de no experimental, transversal de tipo correlacional, con una muestra de 39,840 personas de 15 años de edad a más años, encuestadas por el INEI en el periodo 2017, que luego de la limpieza de los datos, se trabajó con una muestra final de 37, 820.

En el capítulo 5, se muestran los procedimientos paso a paso para la aplicación de las redes neuronales artificiales, 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.

## **CAPITULO I: EL PROBLEMA**

## 1.1. SELECCIÓN DEL PROBLEMA

La investigación en el ámbito de las ciencias sociales y del comportamiento está requiriendo la utilización de técnicas propias de análisis multivariado que, utilizando la segmentación jerárquica basada en el algoritmo CHAID, permitan buscar y describir grupos poblacionales que respondan a determinados rasgos comunes, así como, explicar y pronosticar las causas de determinados fenómenos, comportamientos y situaciones con el fin de conducir programas de intervención (Sanz E. y Ponce de León, A., 2010).

El análisis multivariable facilita la realización de estudios explicativos orientados a estudiar por qué sucede un evento, cómo han llegado a ser los fenómenos lo que son, lo cual permite un tratamiento global y simultáneo de la información. (Vivanco, 1999).

La técnica de segmentación basada en árboles CHAID (Detección Automática de Interacción basada en el Chi Cuadrado, o Chi-Squared Automatic Interaction Detection, por sus siglas en inglés) ha mostrado ser útil para obtener segmentos significativos que sean predictivos de una variable criterio de K categorías (nominal u ordinal). CHAID fue diseñado para detectar, de manera automática, la interacción entre varios predictores categóricos u ordinales para explicar una respuesta categórica, esto se debe al hecho de que CHAID es un algoritmo de selección hacia adelante basado en conteos marginales. (Galindo, M.; Vicente, J.; Dorado, A.; Vicente, P. y Patino, M., 2009)

Cuando se trata de construir un modelo con finalidad predictiva es necesario partir de un conjunto de entrenamiento lo más amplio posible de casos problema con solución conocida. Idealmente, antes de aplicar el modelo debe validarse con otro conjunto de casos (grupo de validación) previamente reservado para esta finalidad y que no haya

sido utilizado en el entrenamiento del modelo. Si la validación es satisfactoria, se demostrará la capacidad de generalización del modelo. Los modelos de árboles de clasificación han despertado un interés creciente en muchos campos de estudio, ya que están constituidos por un conjunto de reglas de interpretación inmediata y de fácil manejo. (Esquerda, A. y Trujillano, J., 2010)

Por otro lado, el Instituto Nacional de Estadística e Informática (INEI), órgano rector del Sistema Nacional de Estadística e Informática (SNEI), ejecuta la “Encuesta Nacional Especializada Sobre Victimización”, en el área urbana de las 24 regiones del país y la Provincia Constitucional del Callao, específicamente en 28 ciudades priorizadas: Arequipa, Trujillo, Ayacucho, Cajamarca, Chiclayo, Chimbote, Cusco, Huancayo, Huánuco, Ica, Iquitos, Juliaca, Piura, Pucallpa, Puno, Tacna, Tarapoto, Abancay, Huaraz, Moquegua, Pasco, Puerto Maldonado, Tumbes, Chachapoyas, Huancavelica, Moyobamba, Área Metropolitana de Lima y Callao. La Encuesta, constituye el medio estadístico más importante y oportuno que dispondrá el país para la obtención de información estadística proveniente de los hogares, que permitirá medir el nivel de victimización, el análisis y diseño de políticas públicas en el tema de seguridad.

Según un último informe técnico de Seguridad Ciudadana proporcionado por el Instituto Nacional de Estadística e Informática (INEI), 26 de cada 100 peruanos de 15 años a más años, fueron víctimas de un hecho delictivo entre los meses de marzo y agosto de este año 2018. La información revela además que, en ciudades de 20 mil a más habitantes, un total de 28.6% personas dijo haber sido víctima de un delito. En tanto, en Lima Metropolitana esta cifra ascendería a 31.0% y 19.6% en los centros poblados del país. En relación a porcentajes detallados, el INEI informó que 13 de

cada 100 habitantes peruanos sufrieron el robo de dinero, cartera o celular. En las ciudades de 20 mil a más habitantes, la cifra fue de 14 de cada 100; en Lima Metropolitana, 15 de cada 100 y en los centros poblados, 9 de cada 100 habitantes. La mayor cifra de victimización se dio en Junín de 42.0%. De acuerdo al sexo, más hombres que mujeres se convirtieron en víctimas del amigo de lo ajeno. Siendo la cifra correspondiente al 27% y 25.4% respectivamente. Por edad, el 34.3% de la población de 15 a 29 años manifestó haber sufrido algún hecho delictivo; seguido por el 28.0%, de 30 a 44 años; y el 22.6%, de 45 a 64 años. Mientras que, por nivel educativo, los universitarios o técnicos fueron víctimas de un delito en un 31.5%; en tanto los estudiantes de postgrado en un 26% y quienes no poseen nivel educativo, en un 10%. (www.americatv.com.pe).

Por todo lo mencionado anteriormente, se planteó la siguiente pregunta de investigación:

**¿Cuáles son los predictores de las víctimas de robo del Perú mediante el algoritmo CHAID, según la Encuesta Nacional Especializada sobre Victimización 2017?**

## **1.2. JUSTIFICACION DEL ESTUDIO**

La principal contribución del presente trabajo de tesis es de índole metodológica, ya que se aplicó una herramienta estadística que no requiere hipótesis inicial sobre la distribución de las variables, aplicando una técnica de minería de datos de árboles de clasificación basada en algoritmos de inducción de reglas conocida como CHAID. El estudio ofreció cuantitativamente los predictores o perfiles de las víctimas de robo del Perú. Así mismo, su utilidad radica en que las autoridades pertinentes del Perú

propongan políticas adecuadas respecto a la seguridad ciudadana, en beneficio de la población.

Por otro lado, es de gran importancia para los estudiantes de la escuela profesional de Estadística e Informática, pues pueden contar con una guía metodológica en la aplicación de los algoritmos CHAID a conjuntos de datos diversos.

### **1.3. OBJETIVOS DE LA INVESTIGACION**

#### **1.3.1. OBJETIVO GENERAL**

Conocer los predictores de las víctimas de robo del Perú mediante el algoritmo CHAID, según la Encuesta Nacional Especializada sobre Victimización 2017.

#### **1.3.2. OBJETIVOS ESPECIFICOS**

- Identificar las variables predictoras de las víctimas de robo del Perú mediante el algoritmo CHAID, según la Encuesta Nacional Especializada sobre Victimización 2017.
- Determinar relaciones bivariadas entre las variables predictoras de las víctimas de robo del Perú.
- Aplicar el Algoritmo CHAID a los predictores de las víctimas de robo del Perú, según la Encuesta Nacional Especializada sobre Victimización 2017.

## **CAPITULO II: MARCO TEORICO CONCEPTUAL**



## 2.1. ANTECEDENTES DEL PROBLEMA

**Rayo & Cortes (2010)**, en su trabajo de investigación titulado “Aplicación del Chaid para identificar las características económico financieras de las empresas inmobiliarias más rentables en España” desarrollado en España, concluyen que, para medir la rentabilidad financiera, se analizan un conjunto de variables independientes relativas a solvencia, liquidez, actividad, rotaciones, equilibrio financiero y estructura económica. La principal contribución es de índole metodológica, ya que se aplican herramientas estadísticas que no requieren hipótesis iniciales sobre la distribución de las variables, aplicando una técnica de minería de datos de árboles de clasificación y regresión basada en algoritmos de inducción de reglas conocida como CHAID. El estudio ofrece cuantitativamente perfiles de éxito definidos con un conjunto de reglas que indican los patrones de las empresas más rentables.

**Moreno, A.; Vicente, P. & Galindo, P. (2016)**, en su trabajo de investigación titulado “Aprendizaje basado en árboles de decisión: un estudio crítico desde Weka, RapidMiner y SPSS Modeler” desarrollado en Ecuador, concluyen que el Data Mining engloba un conjunto de técnicas, que buscan descubrir las relaciones existentes en una gran cantidad de datos. En los últimos años se ha producido un interés generalizado y creciente, motivado por la mayor competitividad en el mercado y por la explosión de datos (Big Data). El predecir el comportamiento humano permite mejorar las ventas, la sanidad, combatir el riesgo financiero, etc. Las técnicas proceden del aprendizaje automático y la estadística, y se clasifican en técnicas de clasificación, segmentación y asociación. Las de clasificación sirven para pronosticar un resultado conocido, por ejemplo, si un cliente comprará o no un producto. Las de asociación son más útiles cuando se desea pronosticar varios resultados conocidos, por ejemplo, los clientes que adquirieron el producto X

también adquirieron Y y Z. Y las de segmentación se usan cuando se desconoce el resultado específico (por ejemplo, identificar nuevos patrones de fraude) por lo que se enfocan en crear grupos similares y describirlos. En las técnicas de clasificación, el objetivo es predecir un evento (valor específico de una variable categórica) o estimar valores de una variable continua. Se trabaja sobre datos históricos donde se tiene un resultado conocido, donde las filas se corresponden a los individuos-instancias a clasificar, y las columnas son los atributos y la etiqueta (variables predictoras y variable objetivo). Los modelos se construyen sobre datos de entrenamiento, y se validan sobre datos de prueba. Aunque entre los estadísticos es más conocido el método CART, el método CHAID es más utilizado en estudios de marketing, sin embargo, tiene importantes limitaciones ya que puede no capturar todas las interacciones entre las variables predictoras y la objetivo.

**Escobar, M. (1998)** en su trabajo de investigación titulado “Las aplicaciones del análisis de segmentación: el procedimiento CHAID” desarrollado en España, concluye que el análisis de segmentación es una técnica de análisis de datos basada en la dependencia entre variables, cuya finalidad es la de formar grupos, configurados con valores de las variables independientes, que sean muy distintos entre sí en la variable dependiente. La lógica de su procedimiento se sustenta en los siguientes pasos: a) agrupación de categorías de los pronosticadores, b) selección de los mejores pronosticadores y c) sucesivas segmentaciones, hasta alcanzar unos límites definidos por los denominados filtros, sobre los grupos formados a partir de los pasos anteriores. Uno de los algoritmos más útiles para sociólogos es el basado en el estadístico Chi-Cuadrado, pues es especialmente indicado para variables dependientes nominales. La utilidad del análisis de segmentación es múltiple. Está especialmente diseñado para propósitos descriptivos, exploratorios e

incluso pronosticadores. Además, con ciertas cautelas, también puede ser útil para un previo análisis causal de las variables.

**Torres, C. (2015)**, en su trabajo de tesis titulado “Métodos de investigación del mercado de créditos de consumo, mediciones de los patrones determinantes según su destino aplicado a instituciones financieras”, para optar el grado de Maestro en Administración de Empresas de la Universidad del Pacífico, Guayaquil – Colombia, concluye que el diagrama de árbol utilizando el método CHAID para bancos pequeños y grandes, facilitó la determinación y establecimiento de los perfiles de los clientes del sistema financiero.

**Ramos, S. (2015)**, en su trabajo de tesis titulado “TAID versus CHAID Búsqueda de perfiles de mujeres trabajadoras en el servicio doméstico” para optar el grado de maestro en Análisis Avanzado de Datos Multivariantes de la Universidad de Salamanca – España, concluye que a pesar de que el algoritmo CHAID presenta importantes limitaciones, entre las que cabe destacar que trabaja con tablas colapsadas y no captura el papel asimétrico de las variables en estudio, sigue siendo el método más utilizado en la bibliografía, en todas las ramas de la Ciencia. Se han encontrado importantes diferencias en la definición de los perfiles de mujeres trabajadoras que se dedican a la actividad del servicio doméstico, al aplicar el algoritmo CHAID y el algoritmo TAID, lo cual pone de manifiesto la necesidad de extender los nuevos métodos de segmentación en los diferentes campos de la ciencia y de desarrollar software específico que contribuya a su divulgación y uso.

**Díaz-Pérez & Bethencourt-Cejas (2016)**, en su trabajo de investigación titulado “CHAID algorithm as an appropriate analytical method for tourism market segmentation” desarrollado en España, concluyen que lo encontrado contrasta fuertemente con la prueba CHAID, ya que se basa en el uso iterativo de la estadística Chi cuadrado, la prueba

estadística no paramétrica por excelencia, donde no se asume ningún tipo de distribución con respecto a las variables independientes involucradas en el análisis. Al utilizar el análisis CHAID, emerge un mayor número de segmentos y variables importantes. Sin lugar a dudas, este último resultado muestra la superioridad de CHAID como método de segmentación aplicable a los mercados turísticos, permitiendo decisiones más precisas y correctas para todos los niveles de la segmentación real de los mercados. Por lo tanto, es posible inferir que los resultados del análisis basado en CHAID coinciden con la naturaleza del problema estudiado mejor que los proporcionados por el análisis discriminante. Por todo lo anterior, Chi-Cuadrado es la estadística básica involucrada en CHAID, que puede discernir entre segmentos (de naturaleza cualitativa) de una manera natural y más poderosa que otros modelos de análisis estadísticos alternativos, ya que Chi-Cuadrado está diseñado para discretos y variables de libre distribución.

**Trujillano, J.; y otros. (2008)**, en su trabajo de investigación titulado “Aproximación a la metodología basada en árboles de decisión (CART). Mortalidad hospitalaria del infarto agudo de miocardio” desarrollado en España, concluyen afirmando que los modelos basados en árboles de decisión ofrecen como ventaja una simplicidad en su utilización e interpretación ya que sus reglas de decisión no necesitan procesos matemáticos para ser interpretadas.

**Navarro, J. & Pastor, E. (2017)**, en su trabajo de investigación titulado “Factores dinámicos en el comportamiento de delincuentes juveniles con perfil de ajuste social. Un estudio de reincidencia”, desarrollado en España, concluyen que la delincuencia juvenil constituye actualmente un fenómeno de naturaleza global, que excede los marcos físico-territoriales del problema. Hasta hace solo una década eran casos puntuales los adolescentes con perfil de ajuste social que llegaban a los juzgados de menores; sin

embargo, en la actualidad se ha convertido en una constante. En definitiva, destacamos la existencia de elementos de riesgo que son determinantes con adolescentes procedentes de entornos de exclusión; sin embargo, no lo son tanto con adolescentes en conflicto con la ley (ACL) con perfil de ajuste, ya que los factores dinámicos son reconducibles. Las fuentes de protección son más consistentes y esto redundará en un contacto con el riesgo intermitente que protege al ACL con perfil de ajuste. Los factores estáticos en adolescentes de este perfil no responden a vivencias especialmente traumáticas y esto produce un desarrollo afectivo positivo. Por otro lado, las redes de soporte social y familiar son fuertes, a pesar de las vicisitudes y experiencias compartidas entre el adolescente y su familia, factor que prioriza la protección mediante redes y mecanismos de inclusión y control social frente a la vulnerabilidad. Por tanto, los factores dinámicos para este perfil permiten canalizar la inclusión mediante pautas prosociales. Los ACL con perfil de ajuste que hemos presentado gozan de conexiones y apoyos, traduciéndose en bajas tasas de evaluación de riesgo y reincidencia delictiva.

**Sánchez, O. & Fuentes, C. (2015)** en su trabajo de investigación titulado “El robo de vehículos y su relación espacial con el contexto sociodemográfico en tres delegaciones centrales de la Ciudad de México (2010)” desarrollado en México, concluyen que los resultados identifican cuatro variables tales como el uso del suelo no residencial, la densidad de población, los hogares con jefatura femenina y concentración de población joven como factores que influyen en el robo de vehículos. Lo anterior permite diseñar políticas públicas de seguridad centradas en los lugares en los que se puedan poner en marcha programas orientados a la prevención situacional del delito, así como la prevención social del delito, orientados a la reconstrucción del tejido social y organización comunitaria.

**Quinteros, D. (2014)** en su trabajo de investigación titulado “Delitos del espacio público y el problema de la “cifra negra”: una aproximación a la no-denuncia en Chile” desarrollado en Chile, concluye que las características más relevantes para comprender la “cifra negra” –que alcanza al 73,1% a nivel general– de los delitos ocurridos en el espacio público contra las personas, esto es, robo con violencia o intimidación, robo por sorpresa y hurto, los que afectan aproximadamente al 20% de la población en Chile cada año. Tras analizar la “cifra negra” con un modelo de regresión logística, según variables de i) victimización, inseguridad y reacción al delito; ii) confianza en las instituciones; y iii) otras de carácter socio-demográfico, se concluye acerca de cuáles –y en qué magnitud–, son los grupos con mayores chances de no denunciar los delitos de los cuáles ha sido víctima, para finalmente relevar, en un segundo nivel, la importancia de generar iniciativas públicas que busquen potenciar la denuncia, en el entendido que permite articular mejores estrategias para la prevención, el control y la sanción de los mismos, junto con entregar una mayor y mejor atención a quienes han sufrido alguno de estos delitos.

**Chaux, E.; y otros. (2013)**, en su trabajo de investigación titulado “Actitudes y dinámicas de robo en un colegio de nivel socioeconómico medio-alto” desarrollado en Colombia, concluyen que los robos son muy frecuentes en ambos contextos donde se involucra armas y amenazas de violencia, pero sus dinámicas difieren radicalmente. Las diferencias pueden estar relacionadas con que en un contexto hay un alto nivel de contacto con pandillas juveniles, que involucran a estudiantes en actividades delictivas desde muy temprano en sus vidas, mientras que en el otro no parece haber contacto con este tipo de grupos juveniles. Sin embargo, también pueden estar asociados otros factores, como posibles diferencias en el nivel de supervisión y aplicación de normas por parte de los profesores, o de los padres de familia, o diferencias en la valoración o rechazo frente a

actos delictivos en diferentes contextos escolares, familiares o comunitarios. En términos generales, el robo es visto por los estudiantes como algo incorrecto. Sin embargo, existe un matiz de grises, dentro de los cuales se clasifican y justifican estos actos. Es decir, las actitudes y posiciones con respecto al robo varían, dependiendo del tamaño y el valor del objeto robado, al igual que el mecanismo o estrategia utilizada para cometerlo. Así, el robo de objetos pequeños y de uso diario, como borradores, lápices y esferos, es percibido como un acto de poca gravedad, como algo común, particularmente cuando estos son encontrados en los pasillos, en los salones y en los pupitres. En cambio, se observó una actitud de rechazo e indignación frente al robo de objetos de valor (como calculadoras, ropa, juguetes y aparatos tecnológicos) y ante los robos en los que el ladrón traspasa los espacios privados del estudiante, como las maletas y los casilleros.

**Cárdenas, O. (2012)** en su trabajo de investigación titulado “¿Con qué probabilidad me toca ser víctima de un delito?” desarrollado en México, concluye que los resultados indican que los atributos personales, como sexo y edad, no están correlacionados con la probabilidad de que una persona sea victimizada. También se encontraron resultados contra intuitivos de acuerdo con la teoría de actividades rutinarias. Esto es, que personas que tienden a evitar ciertas rutas son más propensas a convertirse en víctimas de la delincuencia. Adicionalmente, el tamaño del coeficiente asociado con la variable de nivel académico es de signo positivo y crece conforme el nivel académico representa una mayor escolaridad, sin embargo, sólo los coeficientes asociados con educación media superior y superior son significativos. Vivir en una zona donde se venden drogas reduce la probabilidad de ser victimado. Por otra parte, la presencia de pandillas agranda las probabilidades de caer víctima del delito, mientras que la variable de jornada laboral incrementa la probabilidad de que alguien sea victimizado. De manera particular se

encontró que mujeres jóvenes con estudios de licenciatura, que no trabajan y que viven en una zona donde no se ha detectado venta de drogas, tienen altas probabilidades de convertirse en víctimas del crimen, especialmente si tienen poco tiempo de vivir en la ciudad.

**El Instituto Nacional de Estadística y Geografía de México (2015)**, en su publicación “En Números, Documentos de Análisis y Estadísticas” considera que la delincuencia vulnera las libertades y los derechos fundamentales de las personas, generando condiciones que pueden afectar física, mental, emocional y económicamente a los individuos, y a la sociedad en su conjunto. Reflejo de ello es que la inseguridad se ha vuelto la principal preocupación de la población. Sin embargo, la delincuencia no afecta por igual a las personas, pues existen condiciones que hacen a ciertos grupos poblacionales más proclives a ser víctimas de algún delito, sea por su sexo, por la carencia de esquemas de protección, por una mayor exposición a situaciones de riesgo o por el entorno donde vive, por mencionar algunos. Frente a esta situación, además de las intervenciones del Estado, la población ha emprendido en los últimos años cambios en su comportamiento y estilos de vida.

**El Ministerio de Defensa Nacional – Policía Nacional (2016)**, en su estudio criminológico “Hurto a personas y su afectación en la economía y contexto social” desarrollado en Colombia, concluyen que lograron identificar que el hurto a personas está compuesto por una serie de componentes y características que lo dinamizan, como son los aspectos socioeconómicos que enmarcan una economía delictiva sólida debido a que los avances tecnológicos han transformado la revolución industrial, dando cabida a nuevas generaciones y facilitando la calidad de vida de las personas, lo anterior está implícitamente adherido a las actividades delictivas que buscan rentabilidad criminal



frente a las tecnologías como lo son los celulares, relojes, mp3, Tablet, entre otros elementos que se enmarca la gama de objetivos principales en el ámbito delictivo.

## **2.2. PLANTEAMIENTO TEORICO ATINGENTE**

### **2.2.1. BASES TEORICAS**

#### **A. ALGORITMO CHAID**

Detección automática de interacción de Chi-Cuadrado CHAID (Chi-Square Automatic Interaction Detector), es un tipo de árbol de decisiones. Utiliza la detección automática de interacción de chi-cuadrado (CHAID) para clasificar registros en las categorías del campo objetivo. Las predicciones se basan en combinaciones de valores en los campos de entrada (IBM Knowledge Center). Un árbol de clasificación de CHAID calcula la categoría de destino predicha para cada nodo del árbol. Este tipo de árbol se genera cuando el campo objetivo es categórico. Las estadísticas de rama incluyen las estadísticas de prueba de Chi-cuadrado que han generado una partición. Si la prueba de Chi-cuadrado es significativa, se crea una nueva partición (nodo hijo). Esta partición es estadísticamente diferente del nodo padre. El tamaño del efecto puede proporcionar alguna información sobre el alcance de esta diferencia. Las pruebas de Chi-cuadrado también se utilizan para fusionar categorías en nodos únicos. Si los recuentos para categorías no son significativamente diferentes, se fusionan en un nodo

Dentro del Reconocimiento de Patrones uno de los problemas más estudiados es el de Clasificación Supervisada, en donde se conoce que un universo de objetos se agrupa en un número dado de clases de las cuales se tiene de cada una, una muestra de objetos que se sabe pertenecen a ella y el problema consiste en dado un nuevo objeto poder establecer sus relaciones con cada una de dichas clases. Los algoritmos

de clasificación supervisada tienen como objetivo determinar la pertenencia de un objeto (descrito por un conjunto de atributos) a una o varias clases, basándose en la información contenida en un conjunto de objetos previamente clasificados (conjunto de entrenamiento - CE). Dentro de los algoritmos utilizados para resolver problemas de clasificación supervisada se encuentran los árboles de decisión. Un árbol de decisión es una estructura que se compone de nodos (internos y hojas) y de arcos. Sus nodos internos están caracterizados por uno o varios atributos de prueba y de estos nodos se desprenden uno o más arcos. Cada uno de estos arcos tiene asociado un valor del atributo de prueba y estos valores determinan qué camino seguir en el recorrido del árbol. Los nodos hoja contienen información que permite determinar la pertenencia del objeto a una clase. Las características principales de un árbol de decisión son: construcción sencilla, no necesita determinar de antemano parámetros para su construcción, puede tratar problemas multi-clase de la misma forma en que trabaja con problemas de dos clases, facilidad para ser representado mediante un conjunto de reglas y la fácil interpretación de su estructura (Franco, A.; Carrasco, J.; Sánchez, G. & Martínez, J., 2008).

Los árboles de decisión se construyen mediante un algoritmo conocido como segmentación recursiva, que es el proceso paso a paso para dicha construcción. Los análisis de clasificación basados en árboles de decisión son técnicas de explotación de datos (data mining) que consisten en estudiar grandes masas de datos con el fin de descubrir patrones no triviales. Los patrones no triviales que se estudiarán habitualmente serán los predictivos y los explicativos. Un árbol de decisión representa una serie de pautas basadas en ciertas variables explicativas que se muestran según recorremos el árbol (Rojo, J., 2006). Estos árboles se construyen

mediante un algoritmo que va dividiendo los registros de la base de datos (casos u observaciones) en nodos de forma recursiva, de manera que con cada subdivisión las frecuencias relativas de las categorías de la variable dependiente vayan tendiendo a 0 o a 1.

El análisis de segmentación fue concebido y debe ser utilizado principalmente con una finalidad exploratoria. La razón radica en que su mecanismo consiste en la búsqueda de las mejores asociaciones de las variables independientes con la dependiente. Su potencia, al mismo tiempo que su peligro, reside en la selección automática de aquellas categorías que pronostican mejor los valores de la variable considerada objetivo. Además, segmentar significa dividir y, en consecuencia, permite que se hallen grupos muy distintos en un determinado aspecto. De este modo, las muestras quedan fragmentadas en distintos tipos de personas u objetos cuya descripción constituye un objetivo adicional de esta técnica (Escobar, M., 1998). Esta técnica, desarrollada por Cellard et al. (1967), Bourouche y Tennenhaus (1972), Kass (1980) y Magidson (1989, 1993 y 1993), quien la ha adaptado para el SPSS, tiene como principal característica distintiva de otros algoritmos de segmentación el que la muestra no se segmente de modo binario, o dicho de otro modo, el que se pueden formar segmentos con más de dos categorías al unísono.

Si  $Y$  es una variable respuesta y las  $p$  variables predictoras son  $x_1, x_2, \dots, x_p$ , donde las  $x$  son tomadas fijas y  $Y$  es una variable aleatoria, el problema estadístico es establecer una relación entre  $Y$  y las  $x$  de tal forma que sea posible predecir  $Y$  basado en los valores de  $x$ . Matemáticamente, se quiere estudiar la probabilidad condicional de la variable aleatoria  $Y$ .  $P[Y/x_1, x_2, \dots, x_p]$  o una función de su probabilidad tal como la esperanza condicional.  $E[Y/x_1, x_2, \dots, x_p]$ .

### *Elementos del árbol*

El árbol de la ilustración tiene tres niveles de nodos, el primer nivel tiene un único nodo en la cima llamado nodo raíz. Un nodo interno en el segundo nivel, y tres nodos terminales que están respectivamente en el segundo y tercer nivel. El nodo raíz y el nodo interno son particionados cada uno en dos nodos en el siguiente nivel los cuales son llamados nodos hijos (o ramas) izquierdo y derecho. La completa homogeneidad de los nodos terminales es un ideal raramente alcanzado en el análisis de datos real. De esta manera, el objetivo del algoritmo de segmentación recursiva es hacer las variables resultantes en los nodos terminales tan homogéneas como sea posible. Una medida cuantitativa de la homogeneidad es la noción de impureza.

*División de un nodo:* Para dividir el nodo raíz en dos nodos homogéneos, se debe seleccionar entre los rangos de todas las variables predictoras el valor de la división que más se acerque al límite pureza para cada nodo hijo. La idea es que si el nodo A se divide en AL (rama izquierda) y AR (rama derecha), la pureza de los dos nodos hijos debe ser mayor que la del nodo A. O su impureza menor. Impureza que suele medirse por la mínima probabilidad, la entropía o el índice de Gini.

*Nodos terminales:* El proceso de segmentación recursiva continúa hasta que el árbol sea saturado en el sentido de que los sujetos en los nodos descendientes no se pueden partir en una división adicional. El número total de divisiones permitidas para un nodo disminuye cuando aumentan los niveles del árbol. Cualquier nodo que no pueda o no sea dividido es un nodo terminal.

*Impureza del nodo:* Sea Y una variable dicotómica con valores 0 y 1, el nodo t es más impuro cuando su impureza es máxima con  $P[Y = \text{correcto}] = \frac{1}{2}$ . La función

de impureza tiene una forma cóncava y se puede definir como:  $i(t) = \phi(P[Y = \text{correcto}])$ .

### ***La lógica del análisis de segmentación del algoritmo CHAID***

Tradicionalmente, el análisis de segmentación se ha reducido al estudio de variables dependientes cuantitativas, utilizando el algoritmo presentado por Morgan y Sonquist (1963). Aquí, sin embargo, se centrará la atención en una derivación de esta técnica que se distingue por utilizar, en lugar de la suma cuadrática intergrupos, el estadístico  $\chi^2$  para la selección de los mejores pronosticadores. De esta forma, se detendrá esta exposición en aquellos casos con variable dependiente medida en escala nominal.

Los pasos lógicos que deben seguirse para realizar esta tarea son los siguientes:

- a) Preparación de las variables. Tarea del analista, que debe seleccionar una variable dependiente que sea de interés para el análisis y elegir un conjunto de posibles pronosticadores relevantes (variables nominales, ordinales con pocas categorías, preferiblemente menos de diez, o incluso variables cuantitativas convertidas en discretas) que permitan realizar una descripción y pronóstico óptimo de la primera variable.
- b) Agrupación de las categorías de las variables independientes en el caso de que éstas tengan un perfil similar de la variable dependiente.
- c) Primera segmentación, que consiste en la selección de la variable que mejor prediga la variable dependiente.
- d) Segunda segmentación. Para cada segmento formado en el paso anterior, se busca entre las variables cuyos valores han sido previamente agrupados de la misma forma que en el paso b), la que tenga mayor poder pronosticador.

- e) Sucesivas segmentaciones. Se procede de forma similar al paso anterior en cada grupo formado por la segmentación previa.

***Reducción de las categorías más discriminantes de cada pronosticador (Escobar, M., 1998)***

Este primer paso consiste en seleccionar las categorías de las variables pronosticadoras que realmente discriminan a los sujetos en la variable dependiente. Suponiendo que una determinada variable tuviera  $c$  valores, se trata de convertirlos a un número  $k < c$  que reduzca la complejidad de la segmentación sin pérdida sustancial de información.

Se puede optar por tres modalidades de reducción según sean las características de las variables pronosticadoras:

- 1) Variables nominales: Cada valor de la variable pronosticadora puede ser agregado a cualquier otro valor de la misma variable. Sea, por ejemplo, la variable situación ocupacional con los valores «ocupado», «parado» e «inactivo». De cara a la formación de grupos, la categoría «ocupado» podría formar grupo con «parados» y/o «inactivo». La primera categoría es contigua, pero la segunda no lo es. Este procedimiento también se denominaba libre (free).
- 2) Variables ordinales: Un valor de la variable sólo puede ser agregado a otro si es contiguo en la escala. En el procedimiento anterior, la categoría «ocupado» sólo podría unirse en un primer momento con la categoría «parado». Los «inactivos» podrían agregarse con los «parados»; pero no con «ocupados». Este procedimiento también se conoce con la denominación de monótono. Un ejemplo de pronosticador monótono adecuado es el nivel de estudios. Si esta

variable tuviera como valores «primarios», «secundarios» y «universitarios», el procedimiento permitiría la fusión de las categorías primera y segunda o segunda y tercera, y descartaría la posibilidad de formar un grupo compuesto por sujetos con estudios primarios y universitarios.

- 3) Variables ordinales con valores perdidos: Es similar a la opción anterior, pero permite un mayor grado de libertad, por cuanto un valor, generalmente el «no sabe, no contesta», puede agregarse libremente a cualquier grupo. Si la variable nivel de estudios tuviera el valor «Ns/Nc», con este procedimiento, también denominado flotante (float), los sujetos que no contestasen podrían agruparse con cualquiera de las tres categorías establecidas.
- 4) Variables cuantitativas: Las variables cuantitativas para ser utilizadas en el procedimiento CHAID tienen que ser recodificadas en valores discretos y tratadas como si fueran ordinales.

El funcionamiento de formación de grupos de categorías homogéneas se basa en el estadístico  $\chi^2$ . Los pasos son los siguientes:

- 1) Se forman todos los pares posibles de categorías. Esto dependerá de la opción que se haya preferido dar a un determinado pronosticador. Así, en la variable práctica religiosa, que presenta cinco valores, el número posible de pares sería de 10 (combinaciones de 5 elementos tomados de dos en dos). Si se opta por la opción ordinal, los pares posibles son 4 (número de categorías menos una) sin considerar los valores perdidos. Y si se escogiese la opción ordinal con valores perdidos, las posibilidades serían 7 (2 veces el número de categorías menos 3).
- 2) Para cada posible par se calcula el  $\chi^2$  correspondiente a su cruce con la variable dependiente. El par con más bajo  $\chi^2$ , siempre que no sea significativo formará

una nueva categoría de dos valores fusionados. La condición de que no sea significativo es muy importante porque, caso de que lo fuese, indicaría que las dos categorías que se pretenden fusionar no lo pueden hacer, ya que son heterogéneas entre sí en los valores de la variable dependiente y el objetivo es justo lo contrario, asimilar categorías con comportamiento semejante.

- 3) Si se ha fusionado un determinado par de categorías, se procede a realizar nuevas fusiones de los valores del pronosticador, pero esta vez con una categoría menos, pues dos de las antiguas han sido reducidas a una sola.
- 4) El proceso se acaba cuando ya no pueden realizarse más fusiones porque los  $x^{\wedge}$  ofrecen resultados significativos.

De esta forma, como casos extremos, podría suceder que una variable con  $c$  categorías siguiera con  $c$  grupos, en el supuesto de que todos ellos sean diferentes entre sí; o bien, que las categorías tengan valores tan parecidos en la variable dependiente que se queden reducidos a uno solo, con lo que el poder discriminador del pronosticador sería nulo.

Existe un procedimiento que ahorra gran cantidad de cálculos y posee una razonable base lógica. Se trata de limitarse a la obtención de segmentaciones binarias. Esto implica que, sea cual sea el número de categorías de los pronosticadores, se busque la mejor combinación de ellas que genere sólo dos grupos ( $k = 2$ ). En consecuencia, habría que formar todas las posibles combinaciones de dos grupos con las  $c$  categorías y seleccionar aquél con un  $x^{\wedge}$  mayor. Es evidente que, utilizando estos contrastes binarios, el número de posibilidades de agrupación se reduce. En el caso de una variable ordinal con valores «mucho», «bastante», «poco» y «nada», el número de contrastes sería de 7 con la opción libre y de 3 con la monótona. Biggs



et al. (1991) propusieron la fusión continua de pares de valores hasta que sólo quedara una única dicotomía de valores, denominando a tal procedimiento CHAID exhaustivo.

### *Selección de los mejores pronosticadores*

Una vez que para cada pronosticador se ha realizado la combinación oportuna de categorías, el siguiente paso sería la selección de los mejores pronosticadores. Para hacerlo, hay que calcular para cada uno de ellos su correspondiente  $\chi^2$  y comparar las significaciones obtenidas; sin embargo, es conveniente en este proceso modificar la significación de cada pronosticador con el ajuste de Bonferroni, porque la probabilidad de obtención de un resultado significativo aumenta artificialmente con la proliferación de pruebas estadísticas que implica este análisis.

El proceso de segmentación debe ser examinado en sus distintas fases con el objeto de valorar el comportamiento de los pronosticadores alternativos. El problema estriba en que el programa analiza varias variables en cada paso de la segmentación y tiene que elegir, entre ellas, sólo una. Si en una determinada fase existen varios pronosticadores de similar poder de segmentación, el análisis de la elección efectuada puede conducir a interpretaciones precipitadas. Para descubrir la posible existencia de este problema, habrá que prestar atención en cada segmentación a la significación ajustada del  $\chi^2$  de los pronosticadores alternativos.

Por último, para determinar la capacidad pronosticadora de la segmentación en su conjunto, resulta muy útil cruzar la variable dependiente con una nueva variable compuesta, cuyos valores sean las características de cada uno de los grupos terminales formados por la segmentación. Un coeficiente de asociación, como puede ser la V de Cramer, resume el poder de predicción de los segmentos en su

explicación de la variable dependiente. Otro modo de juzgar la bondad de la segmentación consiste en construir una tabla donde se crucen los datos empíricos de la variable dependiente con los que se pronosticarían con el conocimiento del segmento al que pertenece cada individuo. En la tabla de clasificación, quedan distinguidas las cifras de la diagonal, que son aciertos o coincidencias entre la predicción y lo real, de las que están fuera de ellas, que son equivocaciones. La estimación del riesgo se calcula mediante el cociente entre estas últimas frecuencias y el total número de casos.

### *La finalización del proceso de segmentación*

Si no se pusieran límites al proceso de segmentación, este análisis podría producir una gran cantidad de grupos terminales de tamaño muy pequeño que serían difíciles de interpretar. En un caso extremo, con un número elevado de variables y sin restricción alguna, este análisis produciría tantos grupos como individuos tuviese la muestra. En la situación común de una muestra de 1.000 sujetos con 5 pronosticadores de tres categorías cada uno, el número posible de grupos terminales sería de 243 ( $3^5$ ) con un tamaño medio aproximado de cuatro personas ( $1.000/243$ ). Es conveniente, por tanto, poner límites al proceso de segmentación. Existen cuatro tipos de filtros que evitan la continuación de la segmentación: los de significación, los de asociación, los de tamaño y los de nivel.

### *Filtros de significación*

Son los más utilizados en la técnica CHAID de segmentación. Su criterio consiste básicamente en no permitir segmentaciones que no sean estadísticamente significativas. Por omisión, se sobrentiende que los límites de significación se sitúan en el nivel 0,05, que se corresponde con un nivel de confianza del 95%. Estos

filtros pueden ser aplicados en dos de los procesos explicados anteriormente: bien en la agrupación de categorías de una variable (fusión de valores), bien en la selección del mejor pronosticador (segmentación de grupos).

La aplicación en el primer proceso es en realidad un mecanismo indirecto de finalización de la segmentación. Su efecto opera fundamentalmente en la cantidad de categorías de una determinada variable que van a segmentarse. Consiste en determinar la significación mínima para que dos categorías de una variable queden englobadas en el mismo segmento. El valor -SC, significación de las categorías (alpha for merging)- más comúnmente asumido para este parámetro es el de 0,05. Si la significación de la diferencia en la variable dependiente entre dos categorías de la variable independiente es menor que este valor, se permite rechazar la hipótesis nula con un 95% de confianza y, como consecuencia, las dos susodichas categorías quedan separadas y se puede proseguir la segmentación. En cambio, si el valor es superior a 0,05, las categorías se funden, y, si quedan agrupadas todas las categorías de todas las variables, la segmentación se detiene.

Los valores extremos permiten comprender con mayor eficacia el efecto de este mecanismo. Si se escoge el mayor valor posible del parámetro (1,0), entonces, la agrupación o reducción de categorías de las variables se toma imposible y, siempre que haya significación entre pronosticador y variable dependiente, la segmentación formará con una determinada variable tantos grupos como categorías tenga.

Si se hubiese establecido el criterio con un parámetro superior a dicha cifra, la segmentación hubiese sido más frondosa, siguiendo la metáfora de la representación en forma arbórea. En concreto, cambiando el filtro, la primera

subdivisión de la muestra, en lugar de dar lugar a dos grupos, proporciona tres grupos.

Si, en vez de poner el nivel de significación de la agrupación de las categorías en un valor alto, se situara en un valor bajo (por ejemplo,  $4E-4$ ), entonces, en lugar de producirse más subdivisiones entre los grupos, se generarían menos divisiones entre las categorías, con el riesgo añadido de que una determinada variable no funcione como un buen pronosticador. No siempre sucede esto de forma que implique la detención de la segmentación de un grupo. Lo lógico es esperar que una subdivisión de  $c$  categorías se reduzca a un número  $k$ , inferior al producido por un nivel de significación superior. En este caso, como el número inicial de categorías es igual a 2, la reducción implica la obtención de una sola categoría y de esta forma la segmentación no se lleva a cabo.

El otro mecanismo de control de significación, en lugar de operar sobre la agrupación de categorías, afecta a la selección de variables. Este procedimiento es una forma directa de finalizar la segmentación, porque, después de encontrar el pronosticador con menor significación, si no es inferior al límite establecido (generalmente 0,05), es obvio que no habrá otro pronosticador que cumpla también con esta propiedad, por lo que el proceso de división de la muestra termina. Visto desde sus posibilidades extremas, si se establece este parámetro -SV, significación de la variable (alpha for splitting)- en el valor 1,0, la segmentación se producirá por todas las variables existentes; pero si se determina que el parámetro sea 0,0, entonces la segmentación no se produce ni tan siquiera en el primer nivel, pues la significación empírica de un pronosticador, por muy pequeña que sea, siempre es superior a cero.

### *Filtros de asociación*

Cumplen una función análoga a la de los filtros de significación de pronosticadores. Se pueden aplicar a los siguientes coeficientes de asociación: Phi, V de Cramer, Coeficiente de Contingencia, T de Tschruprow u otros. Se trata de detener la segmentación no porque un determinado cruce no obtenga un mínimo de significación, sino porque el coeficiente de asociación elegido no alcance un determinado nivel. Lo que principalmente diferencia a un procedimiento de otro, es el hecho de que el que opera sobre la asociación no es sensible al número de casos sobre los que se trabaja. Por tanto, en valores equiparables de uno y otro, los filtros de asociación son más permisivos en los niveles más bajos de segmentación. Como los de significación son muy sensibles al número de casos, es muy probable que en el tercer o cuarto nivel el análisis no cumpla las condiciones del filtro, porque los segmentos tengan un tamaño reducido. En cambio, los coeficientes de asociación, por el hecho de eliminar la influencia del número de casos, permiten segmentaciones aun en condiciones de escasos sujetos. En este caso, hay mucho menos acuerdo sobre cuál debe ser el valor del filtro. Como regla de experiencia, se consideran adecuados los valores 0,10 ó 0,20. Sin embargo, el programa Answer Tree del SPSS no contempla la posibilidad de utilizarlos para el control de la segmentación. En todo caso, la opción recomendada para el uso de estos filtros es que se utilicen en conjunción con un filtro de significación, de forma que una segmentación que no sea significativa no se lleve a cabo por muy grande que sea su coeficiente de asociación. El caso contrario, que justifica especialmente el uso de estos filtros, también suele suceder. Se trata de relaciones entre variable dependiente y pronosticador muy significativas, pero con un coeficiente de

asociación bajo, que se dan con frecuencia cuando se trabaja con muestras de elevado número de casos.

### *Filtros de tamaño*

Su principal objetivo consiste en evitar que se formen grupos muy pequeños durante el proceso de segmentación, dado el problema que supone la generalización en estos casos. Si, por ejemplo, se segmentara un grupo de 25 personas de las que un 30% es favorable al aborto, se plantearían dos problemas: por un lado, este grupo no sería representativo en sí de la población; por otro, el valor del 30% tampoco sería un estimador muy preciso con un tamaño de muestra tan reducido.

Los filtros de tamaño pueden aplicarse en dos momentos: después de la segmentación ( $N_d$ , child node) y antes de la segmentación ( $N_a$ , parent node). En el primer caso, no se puede formar un grupo si no tiene un número establecido de componentes. En el segundo, la segmentación se detiene en el supuesto de que haya un grupo que haya descendido de un determinado número de individuos. Es obvio que ambos filtros pueden utilizarse al mismo tiempo. Lo que no tiene sentido es que el filtro antes de la segmentación ( $N_a$ ) sea inferior en número al de después ( $N_d$ ), puesto que de esta forma este último no se aplicaría. Sólo tiene razón que  $N_a$  sea superior a  $N_d$ . Como regla general, se recomiendan unos parámetros de 100 para  $N_a$  y 50 para  $N_d$ . Esto implica la no obtención de grupos inferiores a un medio centenar de personas y la no segmentación de conjuntos con menos de cien componentes.

### *Filtros de nivel*

Por último, existe un cuarto tipo de mecanismo de detención de la segmentación. Consiste en arbitrar un nivel ( $N_s$ , depth) máximo de segmentación. Si se establece

este criterio en 0, la segmentación no tendrá lugar; si en 1, sólo se realizará una segmentación; si en 2, dos tandas. Por tanto, por nivel se entiende cada una de las franjas horizontales del árbol invertido. La primera franja horizontal corresponde al total de la muestra, la segunda a la primera segmentación, la tercera a la segunda. Este filtro evita que se formen múltiples segmentaciones en segmentos desproporcionadamente grandes de la muestra. Asimismo, contribuye a simplificar los resultados en la medida en que reduce directamente el número de variables necesarias para predecir la variable dependiente.

### ***Ventajas y desventajas del método CHAID***

Las ventajas son las siguientes:

*Transparencia:* La forma de un árbol CHAID exhaustivo es intuitiva. Esto significa que el usuario puede comprobar la racionalidad del modelo y, si lo cree necesario, modificar el árbol en su arquitectura según su propia experiencia o conocimiento.

*Construcción sencilla:* Los modelos CHAID exhaustivo pueden manejar fácilmente variables independientes tanto categóricas (como el estado civil) como continuas relacionadas (como los ingresos). Concretamente, si las variables independientes son categóricas y con un alto nivel de cardinalidad (que implícitamente "contienen" relaciones) CHAID exhaustivo dará sin dudas un mejor resultado.

*Seguridad:* El CHAID exhaustivo suele proporcionar más seguridad, debido a que su método de combinación de categorías de variables es más minucioso que el análisis CHAID.

Las desventajas son:

*El volumen de datos:* Requiere grandes masas de datos para asegurarse de que la cantidad de observaciones de los nodos hijos sea significativa.

*Las variables continuas han de estar relacionadas:* Las variables independientes continuas, como los ingresos, deben estar relacionadas en grupos a imagen de los datos categóricos antes de ser usadas en CHAID exhaustivo.

*Tiempo:* Tarda más tiempo en realizar los cálculos de las divisiones de los nodos porque realiza un análisis más minucioso.

## **B. PREDICTORES DE LAS VÍCTIMAS DE ROBO**

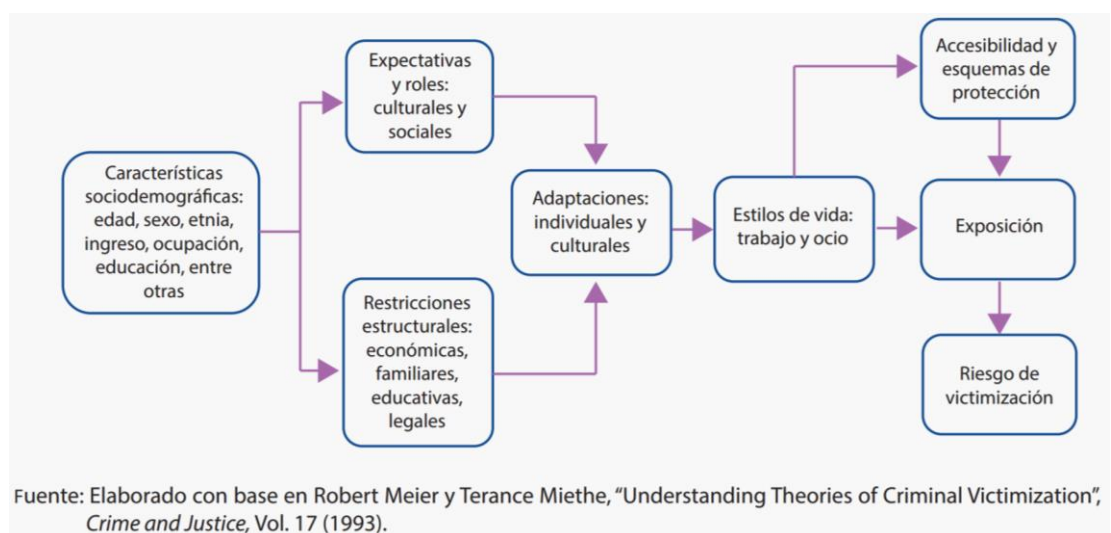
La calidad de vida de las personas depende en buena medida de qué tan seguras se sienten en sus espacios cotidianos. Un ambiente de paz y seguridad es esencial para un pleno ejercicio de los derechos y libertades de las personas.

Desde la perspectiva de la teoría criminológica, la idea subyacente de la victimización es que una conducta delictiva sucederá si un delincuente encuentra las condiciones favorables en tiempo y espacio para perpetrar el delito. A partir de diversos campos del conocimiento se ha buscado identificar a los factores que condicionan la aparición de dichas conductas, factores que ahora sabemos no sólo atañen a los individuos sino también al entorno donde estos se desenvuelven. Diversas teorías del crimen, complementarias entre sí, han resaltado aspectos clave para entender el surgimiento de la delincuencia, y en consecuencia de la victimización en un cierto espacio. Una de ellas pone el énfasis en las manifestaciones de incivildad como condicionante de la inseguridad en las localidades, al señalar que la ausencia de control y los bajos niveles de cohesión social provocan mayor sensación de inseguridad en la comunidad. Otra teoría es la



de vulnerabilidad que postula la existencia de condiciones que hacen más proclives o vulnerables a las personas ante la delincuencia, estas condiciones pueden ser físicas, como la edad y sexo, o de tipo psicosocial como la existencia de lazos familiares positivos que ayudan a prevenir la victimización (Hunter, A., 1978).

Otra de las teorías desarrolladas para explicar la victimización es la de los estilos de vida y las actividades rutinarias (Figura 1), aquí la premisa básica es que existe una serie de características sociodemográficas de los individuos que inciden en su probabilidad de ser víctimas de un delito. Estas características, junto con sus preferencias y gustos, se combinan con elementos culturales, sociales, institucionales y económicos del entorno, lo que modula las percepciones y las expectativas de las personas. Esto significa que la población adapta sus expectativas a su realidad circundante, con lo cual se determina su comportamiento y estilo de vida. Producto de estos cambios, se llevan a cabo por ejemplo mecanismos de protección que pueden ser físicos, como el gasto en instrumentos de seguridad (equipos, pólizas de seguro, etc.), o de tipo social, como el fortalecimiento de las redes familiares y vecinales para prevenir la actividad delincencial (Cohen, L. & Felson, M., 1979).



Otro factor es el grado de exposición y proximidad al delito, el cual conforme un individuo se encuentre más expuesto a situaciones de riesgo o de vulnerabilidad, sea porque está en espacios públicos y en zonas de alta criminalidad o porque su atractivo económico hace más deseables sus bienes a los criminales, se elevará su riesgo de ser víctima (Fabio A.R. Gomes y Lourenco S. Paz, 2008).

Las teorías expuestas anteriormente han servido para estudiar los factores de riesgo de la victimización y eso incluye el estudio de dos conceptos importantes: Víctimas múltiples: son aquellas personas que sufren más de un delito dentro de un periodo determinado; y Víctimas repetidas: son aquellas personas que sufren más de un delito con similitudes en cuanto a su naturaleza o circunstancias. En contraparte, están los entornos de alto riesgo en los que se desenvuelven los individuos, como caminar de noche por la misma calle o viajar en transporte público por rutas inseguras, lo que los hacen más proclives a experimentar de manera repetida un delito (Maureen Outlaw, Barry Ruback y Chester Britt, 2002).

### 2.2.2. DEFINICION DE TERMINOS

**Robo:** El robo es un delito contra el patrimonio, consistente en el apoderamiento de bienes ajenos de otras personas, empleando para ello fuerza en las cosas o bien violencia o intimidación en las personas. Son precisamente estas dos modalidades de ejecución de la conducta las que la diferencia del hurto, que exige únicamente el acto de apoderamiento (<https://es.wikipedia.org/wiki/Robo>).

**Hurto:** Consiste el delito de hurto en el apoderamiento ilegítimo de una cosa mueble ajeno que, a diferencia del robo, es realizado sin fuerza tipificada para acceder o huir del lugar donde se encuentran las cosas, ni personas. La definición del hurto, a diferencia del robo y de la extorsión, requiere siempre apoderamiento, sin usar de formas o modos especiales, como la fuerza para acceder o huir del lugar donde se encuentran las cosas o la violencia física en las personas, características del robo, o como la intimidación para obligar a la entrega, por ejemplo, propia de la extorsión. Se configura el hurto como el tipo básico de apoderamiento (<https://es.wikipedia.org/wiki/Hurto>).

**Distribución Chi-Cuadrado:** En estadística, la distribución de Pearson, llamada también ji cuadrada(o) o chi cuadrado(a) ( $\chi^2$ ), es una distribución de probabilidad continua con un parámetro k que representa los grados de libertad de la variable aleatoria  $X = Z_1^2 + Z_2^2 + \dots + Z_k^2$ , donde  $Z_i$  son variables aleatorias normales independientes de media cero y varianza uno (Córdova, M., 2005).

## **CAPITULO III: HIPOTESIS Y OPERACIONALIZACION DE VARIABLES**

### 3.1. HIPOTESIS

Los predictores de las víctimas de robo del Perú mediante el algoritmo CHAID, según la Encuesta Nacional Especializada sobre Victimización 2017 son los hábitos de la persona y sus características sociodemográficas.

### 3.2. OPERACIONALIZACION DE VARIABLES

Las variables de la presente investigación son:

Cuadro N° 01  
Operacionalización de Variables

<i>Variable</i>	<i>Indicadores</i>	<i>Tipo de variables</i>
Variable independiente: Variables de entrada	Edad	Cuantitativo
	Sexo	Cualitativo
	Nivel educativo	Cualitativo
	Hábitos de la persona	Cualitativo
	Objetos de la persona (cartera, celular, dinero, joyas, etc)	Cualitativo
Variable dependiente: Víctima de robo	Víctima de robo o hurto	Dicotómico

Elaboración: Propia

## **CAPITULO IV: METODOLOGIA**

## **4.1. MATERIALES Y LUGAR DE EJECUCION**

### **4.1.1. LUGAR**

Instituto Nacional de Estadística e Informática.

### **4.1.2. MATERIALES**

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

### **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 tesis.

## **4.2. METODOS**

### **4.2.1. TECNICAS DE RECOLECCION DE DATOS**

Para la presente investigación se recopilará información del sistema de base de datos del INEI (<http://iinei.inei.gob.pe/microdatos/>), de la Encuesta Nacional Especializada sobre Victimización 2017, mediante la aplicación de la minería de datos, mediante la metodología CRISP-DM.

### **4.2.2. TIPO DE ESTUDIO**

- ✓ *Según su propósito o finalidad:* Aplicada
- ✓ *Según su alcance temporal:* Transversal
- ✓ *Según su carácter:* cuantitativa

✓ *Según su nivel de conocimientos que se adquieren o profundidad:* correlacional

✓ *Según el método utilizado:* Analítica e inductiva.

#### **4.2.3. DISEÑO DE LA INVESTIGACION**

La presente investigación es de diseño no experimental, transversal, correlacional.

### **4.3. POBLACION Y MUESTRA**

#### **4.3.1. POBLACION**

La población de estudio comprende todas las personas de 15 y más años de edad residentes habituales de las viviendas particulares del área urbana de las ciudades investigadas. Se excluye del estudio a los miembros de las fuerzas armadas que viven en cuarteles, campamentos, barcos, y otros. También se excluye a las personas que residen en viviendas colectivas (hoteles, hospitales, asilos y claustros religiosos, centros de reclusión penitenciarios, etc.).

#### **4.3.2. MUESTRA**

El tamaño de la muestra nacional para este trabajo es de 37, 820 personas de 15 años de edad a más.

### **4.4. INSTRUMENTOS DE RECOPIACION DE DATOS**

- Técnica:

- Documentaria: Se recopiló a partir de la Encuesta Nacional Especializada sobre Victimización 2017.

- Instrumento:

- Encuesta Nacional Especializada sobre Victimización 2017.



#### **4.5. ANALISIS ESTADISTICO E INTERPRETACION DE LA INFORMACION**

Para el procesamiento y organización de los datos se utilizó el software IBM SPSS Modeler v18 versión trial. 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 las tablas estadísticas. Además, se hizo uso del análisis del Algoritmo CHAID.

## **CAPITULO V: ANALISIS Y DISCUSION DE RESULTADOS**

## 5.1. ANALISIS DE LOS DATOS

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

### **Comprensión del problema**

El problema a solucionar fue la identificación de los predictores de las víctimas de robo del Perú, como parte de la Encuesta Nacional Especializada sobre Victimización 2017.

### **Comprensión de los datos**

Los datos fueron recolectados de la página web del INEI, del repositorio de base de datos, del enlace microdatos (<http://iinei.inei.gob.pe/microdatos/>), de la base de datos de la Encuesta Nacional Especializada sobre Victimización 2017, contando con un total de 22 módulos, una cantidad de registros que oscila entre 59 a 133409 que hacen un total de 1970 variables en total en los 22 módulos.

Tabla N° 01

*Distribución de la cantidad de registros y variables en los módulos de las bases de datos*

#	Código Módulo	Módulo	Cantidad de registros	Cantidad de variables
1	1289	Capítulo 100: Datos de la vivienda	38204	67
2	1290	Capítulo 200: Datos de las personas	133409	37
3	1291	Capítulo 300: Seguridad ciudadana hogar	38204	95
4	1292	Capítulo 300: Seguridad ciudadana persona	38204	188
5	1293	Capítulo 401: Hecho de victimización a la vivienda	38204	62
6	1294	Capítulo 402: Hecho de victimización a la persona	38204	59
7	1295	Capítulo 501: Robo e intento de robo en la vivienda	3637	110
8	1296	Capítulo 502: Robo e intento de robo del vehículo automotor	293	96
9	1297	Capítulo 503: Robo e intento de robo de autopartes del vehículo automotor	1105	106
10	1298	Capítulo 504: Robo e intento de robo de motocicleta, mototaxi y/o moto triciclo	519	96
11	1299	Capítulo 505: Robo e intento de robo de bicicleta y/o triciclo	189	95
12	1300	Capítulo 506: Vandalismo	717	106
13	1301	Capítulo 507: Secuestro	59	100
14	1302	Capítulo 601: Robo e intento de robo con violencia o amenaza	7057	109
15	1303	Capítulo 602: Agresión física y verbal	2828	121
16	1304	Capítulo 603: Amenaza	574	104
17	1305	Capítulo 604: Extorsión	235	102
18	1306	Capítulo 605: Fraude bancario	184	57
19	1307	Capítulo 606: Fraude al consumidor	709	71
20	1308	Capítulo 607: Otras modalidades de estafa y fraude	2302	70
21	1309	Capítulo 700: Tenencia de armas de fuego	38204	49
22	1310	Capítulo 800: Características generales de la persona seleccionada	38204	70

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

Se integraron los módulos y se depuraron las variables, quedando las siguientes variables seleccionadas:

Y : Víctima de robo, es la persona que sufrió robo con violencia o amenaza de sus objetos personales, como cartera, celular, joyas, etc. usando la fuerza o amenaza (C402\_P1\_1), así como, la persona que sufrió robo sin violencia ni amenaza o hurto de sus objetos personales, como cartera, celular, joyas, etc. sin usar la fuerza ni la amenaza (C402\_P1\_3).

X1 : Edad en años cumplidos (C200\_P8\_A)

X2 : Sexo (C200\_P7)

X3 : Percepción de inseguridad en su casa (C300\_P1\_1)

X4 : Percepción de inseguridad caminando por la calle (C300\_P1\_3)

X5 : Percepción de inseguridad caminando en su zona o barrio de noche y solo/a (C300\_P2)

X6 : Percepción de la delincuencia en el país (C300\_P6\_1)

X7 : Cambio de hábito en salir de noche (C300\_P10\_1)

X8 : Cambio de hábito en permitir que sus hijos menores de edad salgan solos (C300\_P10\_2)

X9 : Cambio de hábito en visitar parientes o amigos (C300\_P10\_3)

X10: Cambio de hábito en tomar taxis en la calle (C300\_P10\_4)

X11: Cambio de hábito en usar transporte público (C300\_P10\_5)

- X12: Cambio de hábito en llevar mucho dinero en efectivo (C300\_P10\_6)
- X13: Cambio de hábito en ir a la Institución Educativa, Institutos Tecnológicos, Universidad, etc. (C300\_P10\_7)
- X14: Cambio de hábito en ir al cine o al teatro (C300\_P10\_8)
- X15: Cambio de hábito en salir a caminar (C300\_P10\_9)
- X16: Cambio de hábito en usar joyas (C300\_P10\_10)
- X17: Cambio de hábito en salir a comer afuera (C300\_P10\_11)
- X18: Cambio de hábito en llevar consigo tarjetas de crédito o débito (C300\_P10\_12)
- X19: Cambio de hábito en ir al estadio (C300\_P10\_13)
- X20: Cambio de hábito en frecuentar centros comerciales (C300\_P10\_14)
- X21: Cambio de hábito en viajar por carretera a otra provincia/departamento (C300\_P10\_15)
- X22: Cambio de hábito en dejar la casa sola (C300\_P10\_16)
- X23: Cambio de hábito en llegar muy tarde a casa (C300\_P10\_17)
- X24: Vigilancia de la PNP por la zona o barrio (C300\_P12\_1)
- X25: Vigilancia del serenazgo por la zona o barrio (C300\_P12\_2)
- X26: Vigilancia de la PNP y serenazgo (conjunto) por la zona o barrio (C300\_P12\_3)
- X27: Calificación del trabajo realizado por la comisaria de la zona o barrio (C300\_P14)

X28: Estrato Sociodemográfico (ESTRATO\_SOCIOE)

X29: Nivel de estudios (C800\_P3)

X30: Trabajo remunerado (C800\_P9)

X31: Ocupación principal (C800\_P15)

### **Preparación de los datos**

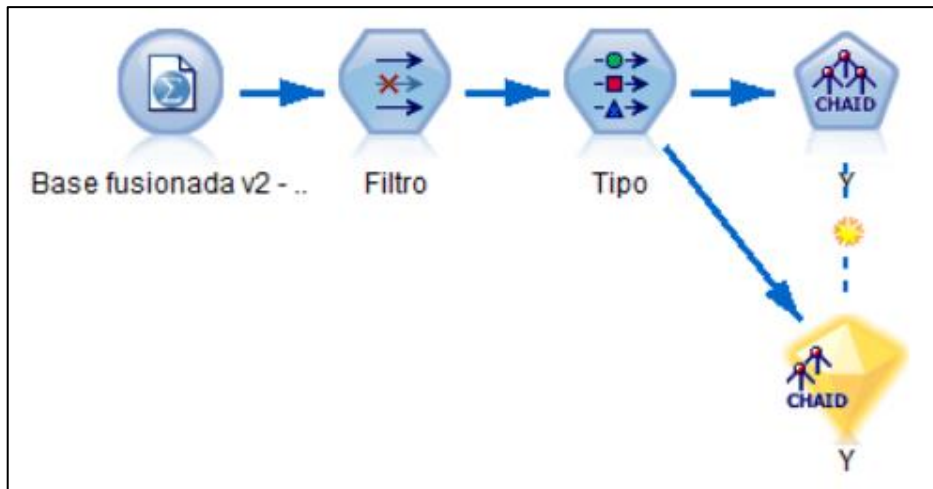
En este punto, se realizó la preparación de los datos, para ello se fusionaron los módulos de los archivos anteriormente mencionados, para su posterior filtro y selección de las variables seleccionadas. Además, se eliminaron los registros con campos en blanco, registros con pérdidas de datos y variables con más del 10% de pérdida de datos.

Después de la limpieza de los datos, la base de datos estuvo conformada por 31 variables de entrada y una variable de salida (anteriormente señalados) y un total de 37, 820 casos.

### **Modelado**

El Algoritmo CHAID aplicado a los predictores de las víctimas de robo del Perú, según la Encuesta Nacional Especializada sobre Victimización 2017, fue obtenido mediante el paquete IBM SPSS Modeler v18.0, para ello se exportó la base de datos del SPSS al Modeler, se realizó el filtro de los datos y luego se tipificó cada una de las variables, para posteriormente aplicar el Algoritmo CHAID a los datos.

Ilustración N° 01



*Ilustración 1.* Modelado del Algoritmo CHAID aplicado a los predictores de las víctimas de robo del Perú, según la Encuesta Nacional Especializada sobre Victimización 2017.

El algoritmo CHAID, presenta la siguiente estructura:

- 10 niveles de profundidad del árbol como máximo
- Alfa para división: 0,05
- Alfa para fusión: 0,05
- Épsilon para convergencia: 0,001
- Número máximo de iteraciones para la convergencia: 100
- Permitir división de categorías fusionadas: true
- Método de chi-cuadrado: Pearson
- Criterios de parada: Utilizar valor absoluto
- Número mínimo de registros en rama padre: 1.000
- Número mínimo de registros en rama hija: 500
- Utilizar costes de clasificación errónea: falso



Tabla N° 02  
*Importancia del predictor de la víctima de robo*

Nodos	Detalle	Importancia
X1	Grupo de edad en riesgo	23,5%
X4	Cómo se siente caminando por la calle	20,8%
X5	Caminar por la zona o barrio sola y de noche	11,2%
X12	Ha dejado de llevar mucho dinero en efectivo	10,7%
X29	Nivel educativo alcanzado	9,5%
X27	Calificación del trabajo de la comisaria	5,1%
X20	Ha dejado de frecuentar centros comerciales	3,7%
X14	Ha dejado de ir al cine o al teatro	3,7%
X2	Sexo	2,5%
X31	Ocupación principal o negocio	2,5%
X13	Ha dejado de ir a la Institución Educativa, Tecnológicos, Universidad, etc.	2,1%
X18	Ha dejado de llevar consigo tarjetas de crédito o débito	1,4%
X7	Ha dejado de salir de noche	1,1%
X10	Ha dejado de tomar taxis en la calle	0,7%
X19	Ha dejado de ir al estadio	0,7%
X30	Trabaja	0,4%
X24	Existe vigilancia de la Policía Nacional del Perú	0,4%

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

De la tabla 2, se observa que las variables predictoras de las víctimas de robo más importante son la edad de riesgo alto (23,5%), seguido por sentirse inseguro al caminar por la calle (20,8%), sentirse inseguro al caminar por la zona o barrio solo(a) y de noche (11,2%), no llevar mucho dinero en efectivo (10,7%) y contar con un nivel educativo superior hasta postgrado (9,5%), entre estas cinco predictoras se acumula un total de 75,8% de probabilidad de ser víctima de robo.

Tabla N° 03  
*Prueba de Chi-Cuadrado de Pearson del Predictor respecto a la Víctima de robo*

Nodos	Categorías	Víctima de robo		
		Chi-cuadrado de Pearson	G.L.	Sig.
X1	0 = Riesgo bajo de robo (15 a 17 y 41 a 99 años) 1 = Riesgo alto de robo (18 a 40 años de edad)	832,79	1	0,000
X4	1 = Inseguro 2 = Seguro	602,04	1	0,000
X5	1 = Inseguro 2 = Seguro	521,72	1	0,000
X12	1 = Si 2 = No 3 = No lleva mucho dinero en efectivo	418,29	2	0,000
X29	0 = Sin nivel hasta secundaria completa 1 = Superior hasta postgrado	432,04	1	0,000
X27	1 = Mala 2 = Buena	204,37	1	0,000
X20	1 = Si 2 = No 3 = No va a los centros comerciales	371,24	2	0,000
X14	1 = Si 2 = No 3 = No va al cine o teatro	342,61	2	0,000
X2	1 = Hombre 2 = Mujer	8,23	1	0,004
X31	1 = Empleador o patrono 2 = Trabajador independiente 3 = Empleado 4 = Obrero 5 = Trabajador familiar no remunerado 6 = Trabajador del hogar 7 = Otros	141,90	6	0,000
X13	1 = Si 2 = No 3 = No va al instituto ni institución educativa	111,59	2	0,000
X18	1 = Si 2 = No 3 = No tiene tarjeta de crédito o débito	319,94	2	0,000
X7	1 = Si 2 = No 3 = No sale de noche	283,40	2	0,000
X10	1 = Si 2 = No 3 = No toma taxi en la calle	428,89	2	0,000
X19	1 = Si 2 = No 3 = No va al estadio	62,00	2	0,000
X30	1 = Si 2 = No	183,83	1	0,000
X24	1 = Si 2 = No 3 = No cuentan con la Policía Nacional del Perú	76,03	2	0,000

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la tabla 3 se observa que todas las variables seleccionadas por el algoritmo CHAID son significativas, por lo tanto, dichas variables se encuentran asociados con la variable víctima de robo.

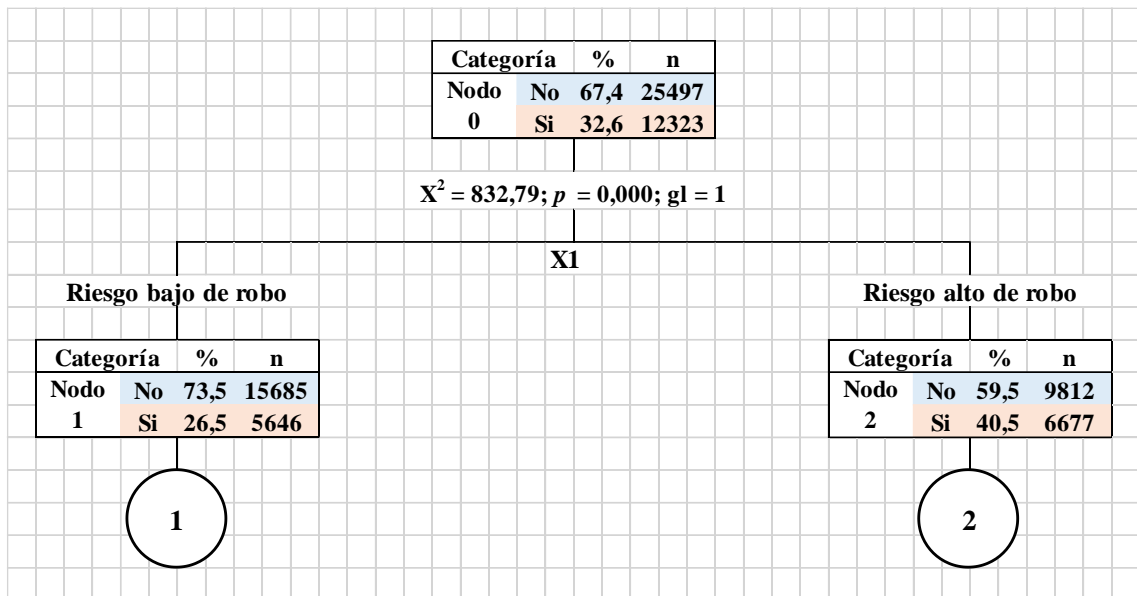


Figura N° 01: Diagrama del árbol mediante el Algoritmo CHAID (Primera parte)

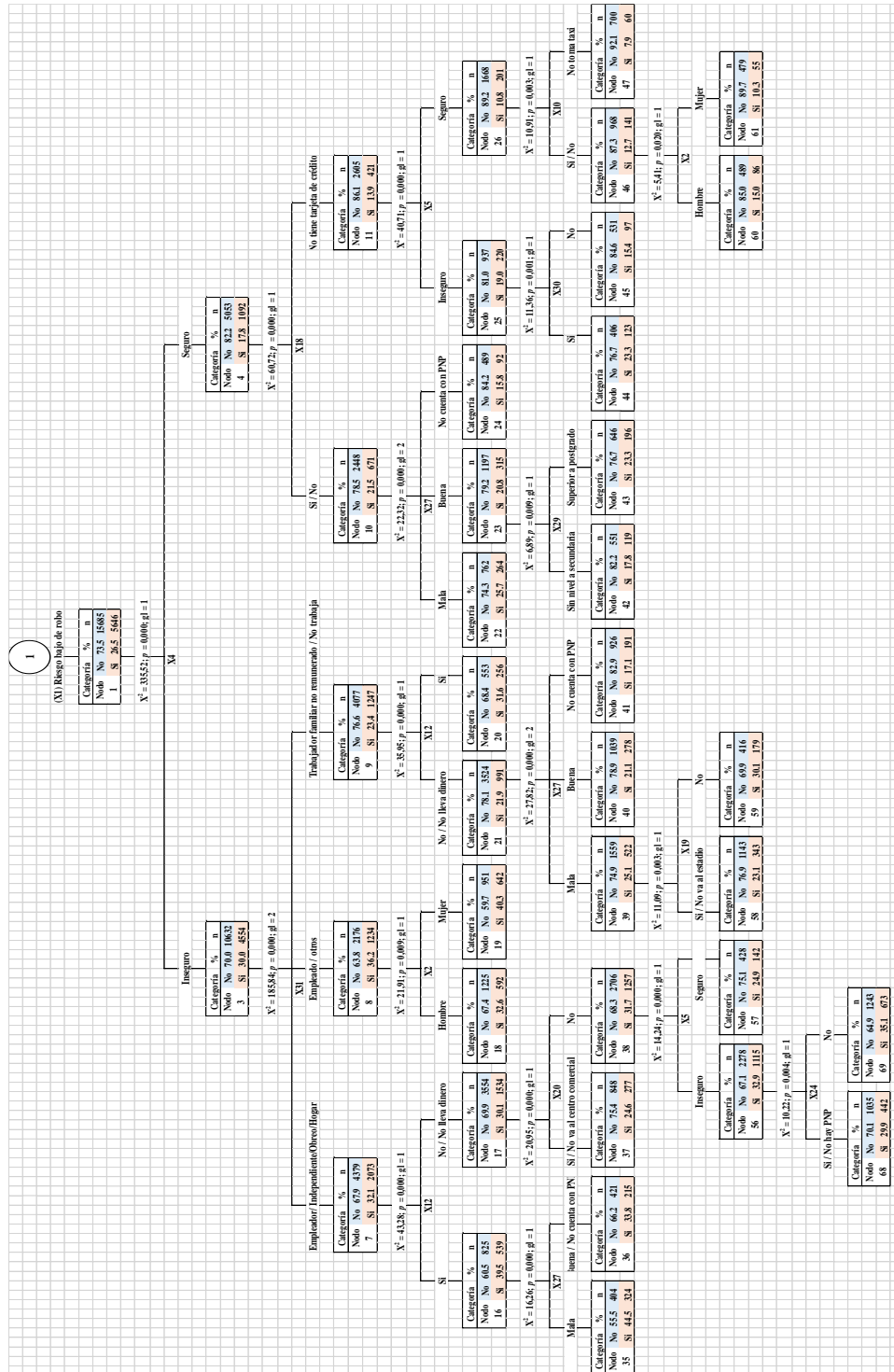


Figura N° 02: Diagrama del árbol mediante el Algoritmo CHAID (Segunda parte)

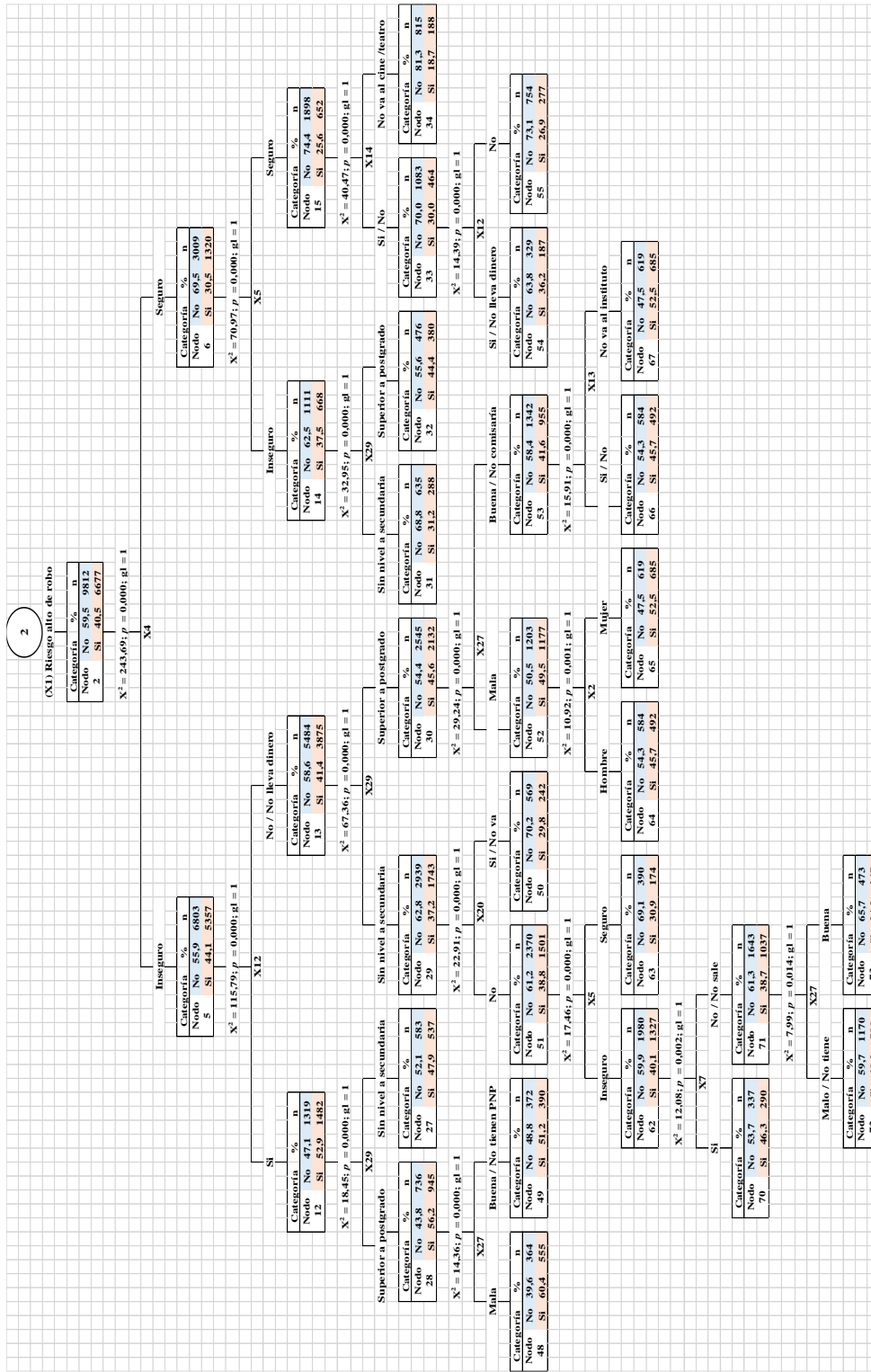


Figura N° 03: Diagrama del árbol mediante el Algoritmo CHAID (Tercera parte)

Los resultados del Algoritmo CHAID en la primera rama (tabla 4) muestra asociación entre la víctima de robo y la edad de la víctima ( $X^2 = 832,79; p < 0,000; gl = 1$ ), con un odds ratio  $OR = 1,890$  (*I. C.* 95%: 1,810 – 1,975), significando que la edad de riesgo alto de robo presenta 1,890 veces de posibilidad de ser víctima de robo que las personas con edad de riesgo bajo de robo. Así mismo, el 54,2% de las víctimas de robo presentan edades en riesgo alto de robo, es decir, son víctimas que tienen entre 18 a 40 años de edad, mientras que el 61,5% de las personas que no fueron víctimas de robo presentan edades en riesgo bajo de robo, es decir, no son víctimas que tienen entre 15 a 17 años y de 41 a 99 años de edad.

Tabla N° 04  
*Víctima de robo según edad de la víctima*

Edad	Víctima de robo			
	No		Si	
	N	%	N	%
Riesgo alto de robo	9812	38,5%	6677	54,2%
Riesgo bajo de robo	15685	61,5%	5646	45,8%
<b>Total</b>	<b>25497</b>	<b>100,0%</b>	<b>12323</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la segunda rama (tabla 5), se muestra asociación entre la víctima de robo con edad de alto riesgo en robo y la forma de sentirse al caminar por la calle de la víctima ( $X^2 = 243,69; p < 0,001; gl = 1$ ), con un odds ratio  $OR = 1,795$  (*I. C.* 95%: 1,667 – 1,933), significando las personas que se sienten inseguros al caminar por la calle presentan 1,795 veces de posibilidad de ser víctima de robo que las personas que se sienten seguros al caminar por la calle. Así mismo, el 80,2% de las víctimas de robo se sienten inseguros al caminar por la calle, mientras que el 30,7% de las personas que no fueron víctimas de robo se sienten seguros al caminar por la calle.

Tabla N° 05  
*Víctima de robo con edad en alto riesgo de robo según forma de sentirse al caminar por la calle*

Forma de sentirse al caminar por la calle	Víctima de robo			
	No		Si	
	N	%	N	%
Segura	3009	30,7%	1320	19,8%
Insegura	6803	69,3%	5357	80,2%
<b>Total</b>	<b>9812</b>	<b>100,0%</b>	<b>6677</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

Tabla N° 06  
*Víctima de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle según ocupación principal o negocio de la víctima*

Ocupación principal o negocio	Víctima de robo			
	No		Si	
	N	%	N	%
Empleador / Independiente /Obrero /Hogar	4379	41,2%	2073	45,5%
Empleado / Otros	2176	20,5%	1234	27,1%
Trabajador no remunerado / No trabaja	4077	38,3%	1247	27,4%
<b>Total</b>	<b>10632</b>	<b>100,0%</b>	<b>4554</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la tercera rama (tabla 6), se muestra asociación entre la víctima de robo con edad de bajo riesgo en robo, forma insegura de sentirse al caminar por la calle con la ocupación principal o negocio de la víctima ( $X^2 = 185,84; p < 0,001; gl = 2$ ), con un odds ratio  $OR = 1,369$  (I.C. 95%: 1,307 – 1,435), significando las personas con ocupación principal de empleados o empleadores presentan 1,369 veces de posibilidad de ser víctima de robo que las personas con trabajos no remunerados o sin empleo. Así mismo, el 27,1% de las víctimas de robo con edad de riesgo bajo de robo, que se sienten inseguros al caminar por la calle son empleados o tienen otras ocupaciones, mientras que el 38,3% de

las personas que no fueron víctimas de robo con edad de riesgo bajo de robo que se sienten inseguros al caminar por la calle son trabajadores familiares no remunerados o simplemente no cuentan con algún tipo de trabajo.

Tabla N° 07  
*Víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle según cambio de hábito al llevar mucho dinero en efectivo*

Ha dejado de llevar mucho dinero en efectivo	Víctima de robo			
	No		Si	
	N	%	N	%
No / No lleva mucho dinero en efectivo	5484	80,6%	3875	72,3%
Si	1319	19,4%	1482	27,7%
<b>Total</b>	<b>6803</b>	<b>100,0%</b>	<b>5357</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

Así mismo (tabla 7), se muestra asociación entre la víctima de robo con edad de alto riesgo en robo, forma insegura de sentirse al caminar por la calle con el cambio de hábito al llevar mucho dinero en efectivo ( $X^2 = 115,79; p < 0,001; gl = 1$ ), con un odds ratio  $OR = 1,590$  (I.C. 95%: 1,461 – 1,731), significando las personas que si han cambiado el hábito de llevar mucho dinero en efectivo presentan 1,590 veces de posibilidad de ser víctima de robo que las personas que no llevan mucho dinero en efectivo o aún no han cambiado su hábito de llevar mucho dinero en efectivo. Así mismo, el 27,7% de las víctimas de robo con edad de riesgo alto de robo, que se sienten inseguros al caminar por la calle si han cambiado el hábito de llevar mucho dinero en efectivo, mientras que el 80,6% de las personas que no fueron víctimas de robo con edad de riesgo alto de robo que se sienten inseguros al caminar por la calle no han cambiado el hábito de llevar mucho dinero en efectivo o simplemente no llevan dinero en efectivo.



Tabla N° 08

*Víctima de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle con ocupación principal de empleado y otras ocupaciones según sexo de la víctima*

Sexo	Víctima de robo			
	No		Si	
	N	%	N	%
Hombre	1225	56,3%	592	48,0%
Mujer	951	43,7%	642	52,0%
<b>Total</b>	<b>2176</b>	<b>100,0%</b>	<b>1234</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la cuarta rama (tabla 8), se muestra asociación entre la víctima de robo con edad de bajo riesgo en robo, forma insegura de sentirse al caminar por la calle con ocupación principal de empleado y otras ocupaciones con el sexo de la víctima ( $X^2 = 21,91; p < 0,009; gl = 1$ ), con un odds ratio  $OR = 1,397$  (I.C. 95%: 1,214 – 1,607), significando que las personas con sexo femenino presentan 1,397 veces de posibilidad de ser víctima de robo que las personas de sexo masculino. Así mismo, el 52,0% de las víctimas de robo con edad de riesgo bajo de robo, que se sienten inseguros al caminar por la calle, de ocupación principal empleado y otras ocupaciones son de sexo femenino, mientras que el 56,3% de las personas que no fueron víctimas de robo con edad de riesgo bajo de robo que se sienten inseguros al caminar por la calle, de ocupación principal empleado y otras ocupaciones son de sexo masculino.

Tabla N° 09

*Víctima de robo con edad en bajo riesgo de robo, forma segura de sentirse al caminar por la calle con cambio de hábito en haber dejado de llevar consigo tarjetas de crédito o débito según calificación del trabajo de la comisaria*

Calificación del trabajo de la comisaria	Víctima de robo			
	No		Si	
	N	%	N	%
Mala	762	31,1%	264	39,3%
Buena	1197	48,9%	315	46,9%
No cuenta con comisaria	489	20,0%	92	13,7%
<b>Total</b>	<b>2448</b>	<b>100,0%</b>	<b>671</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la cuarta rama segunda relación (tabla 9), se muestra asociación entre la víctima de robo con edad de bajo riesgo en robo, forma segura de sentirse al caminar por la calle con cambio de hábito en haber dejado de llevar consigo tarjetas de crédito o débito y la calificación del trabajo de la comisaria ( $X^2 = 22,32; p < 0,001; gl = 2$ ), con un odds ratio  $OR = 1,348$  (I. C. 95%: 1,190 – 1,526), significando que las personas con calificación de mala el trabajo de la comisaria presentan 1,348 veces de posibilidad de ser víctima de robo que las personas con calificación de bueno respecto al trabajo de la comisaria. Así mismo, el 39,3% de las víctimas de robo con edad de riesgo bajo de robo, que se sienten seguros al caminar por la calle, que cuentan con tarjeta de crédito califican como mala el trabajo de la comisaria, mientras que el 48,9% de las personas que no fueron víctimas de robo con edad de riesgo bajo de robo que se sienten seguros al caminar por la calle, que cuentan con tarjeta de crédito califican como buena el trabajo de la comisaria.

Tabla N° 10

*Víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que consideran insegura caminar por la zona o barrio sola y de noche según nivel educativo*

Nivel educativo	Víctima de robo			
	No		Si	
	N	%	N	%
Sin nivel hasta secundaria completa	635	57,2%	288	43,1%
Superior hasta postgrado	476	42,8%	380	56,9%
<b>Total</b>	<b>1111</b>	<b>100,0%</b>	<b>668</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la cuarta rama tercera relación (tabla 10), se muestra asociación entre la víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que consideran insegura caminar por la zona o barrio sola y de noche con el nivel educativo (  $X^2 = 32,95; p < 0,001; gl = 1$  ), con un odds ratio  $OR = 1,760$  (I.C. 95%: 1,450 – 2,137), significando que las personas con nivel educativo superior hasta postgrado presentan 1,760 veces de posibilidad de ser víctima de robo que las personas con nivel educativo de secundaria a sin nivel educativo. Así mismo, el 56,9% de las víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que consideran insegura caminar por la zona o barrio sola y de noche presentan un nivel educativo superior hasta postgrado, mientras que el 57,2% de las personas que no fueron víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que consideran insegura caminar por la zona o barrio sola y de noche presentan un nivel educativo de secundaria a sin nivel educativo.

En la cuarta rama cuarta relación (tabla 11), se muestra asociación entre la víctima de robo edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que han cambiado su hábito de llevar mucho dinero en efectivo y el nivel educativo ( $X^2 = 18,45; p < 0,001; gl = 1$ ), con un odds ratio  $OR = 1,394$  (I.C. 95%: 1,198 – 1,622),

significando que las personas con nivel educativo superior hasta postgrado presentan 1,394 veces de posibilidad de ser víctima de robo que las personas con nivel educativo de secundaria a sin nivel educativo. Así mismo, el 63,8% de las víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que han cambiado su hábito de llevar mucho dinero en efectivo presentan un nivel educativo superior hasta postgrado, mientras que el 44,2% de las personas que no fueron víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que han cambiado su hábito de llevar mucho dinero en efectivo presentan un nivel educativo de secundaria a sin nivel educativo.

Tabla N° 11  
*Víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que han cambiado su hábito de llevar mucho dinero en efectivo según nivel educativo*

Nivel educativo	Víctima de robo			
	No		Si	
	N	%	N	%
Sin nivel hasta secundaria completa	583	44,2%	537	36,2%
Superior hasta postgrado	736	55,8%	945	63,8%
<b>Total</b>	<b>1319</b>	<b>100,0%</b>	<b>1482</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la quinta rama primera relación (tabla 12), se muestra asociación entre la víctima de robo edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador /independiente /obrero /hogar, que ha dejado de llevar mucho dinero en efectivo y la calificación del trabajo de la comisaria ( $X^2 = 16,26; p < 0,001; gl = 1$ ), con un odds ratio  $OR = 1,570$  (*I.C.* 95%: 1,260 – 1,957), significando que las personas con calificación de mala al trabajo de la comisaria presentan 1,570 veces de posibilidad de ser víctima de robo que las personas con calificación de buena al trabajo de la comisaria. Así mismo, el 60,1% de las víctimas de robo con edad

en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador /independiente /obrero /hogar, que ha dejado de llevar mucho dinero en efectivo califican como mala el trabajo de la comisaria, mientras que el 51,0% de las personas que no fueron víctimas de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador /independiente /obrero /hogar, que ha dejado de llevar mucho dinero en efectivo califican como bueno el trabajo de la comisaria.

Tabla N° 12

*Víctima de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador /independiente /obrero /hogar, que ha dejado de llevar mucho dinero en efectivo según calificación del trabajo de la comisaria*

Nivel educativo	Víctima de robo			
	No		Si	
	N	%	N	%
Mala	404	49,0%	324	60,1%
Buena / No cuenta con comisaria	421	51,0%	215	39,9%
<b>Total</b>	<b>825</b>	<b>100,0%</b>	<b>539</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

Tabla 13

*Víctima de robo con edad en bajo riesgo de robo, forma segura de sentirse al caminar por la calle, que no cuentan con tarjeta de crédito o débito, que consideran insegura caminar por la zona o barrio sola y de noche según condición del trabajo*

¿Trabaja?	Víctima de robo			
	No		Si	
	N	%	N	%
Si	406	43,3%	123	55,9%
No	531	56,7%	97	44,1%
<b>Total</b>	<b>937</b>	<b>100,0%</b>	<b>220</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la quinta rama segunda relación (tabla 13), se muestra asociación entre la víctima de robo edad en bajo riesgo de robo, forma segura de sentirse al caminar por la calle, que no

cuentan con tarjeta de crédito o débito, que consideran insegura caminar por la zona o barrio sola y de noche y la condición del trabajo ( $X^2 = 11,36; p < 0,001; gl = 1$ ), con un odds ratio  $OR = 1,658$  (*I.C.* 95%: 1,234 – 2,229), significando que las personas con trabajo presentan 1,658 veces de posibilidad de ser víctima de robo que las personas sin trabajo. Así mismo, el 55,9% de las víctimas de robo con edad en bajo riesgo de robo, forma segura de sentirse al caminar por la calle, que no cuentan con tarjeta de crédito o débito, que consideran insegura caminar por la zona o barrio sola y de noche cuentan con un trabajo, mientras que el 56,7% de las personas que no fueron víctimas de robo con edad en bajo riesgo de robo, forma segura de sentirse al caminar por la calle, que no cuentan con tarjeta de crédito o débito, que consideran insegura caminar por la zona o barrio sola y de noche no cuentan con un trabajo.

Tabla N° 14

*Víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado según calificación del trabajo de la comisaria*

Calificación del trabajo de la comisaria	Víctima de robo			
	No		Si	
	N	%	N	%
Mala	364	49,5%	555	58,7%
Buena / No cuenta con comisaria	372	50,5%	390	41,3%
<b>Total</b>	<b>736</b>	<b>100,0%</b>	<b>945</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la quinta rama tercera relación (tabla 14), se muestra asociación entre la víctima de robo edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado y la calificación del trabajo de la comisaria ( $X^2 = 14,36; p < 0,001; gl = 1$ ), con un

odds ratio  $OR = 1,454$  (*I.C.* 95%: 1,198 – 1,766), significando que las personas que califican como mala el trabajo de la comisaria presentan 1,454 veces de posibilidad de ser víctima de robo que las personas que califican como buena el trabajo de la comisaria. Así mismo, el 58,7% de las víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado califican como mala el trabajo de la comisaria, mientras que el 50,5% de las personas que no fueron víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado califican como buena el trabajo de la comisaria.

Tabla N° 15

*Víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado según calificación del trabajo de la comisaria*

Calificación del trabajo de la comisaria	Víctima de robo			
	No		Si	
	N	%	N	%
Mala	1203	47.3%	1177	55.2%
Buena / No cuenta con comisaria	1342	52.7%	955	44.8%
<b>Total</b>	<b>2545</b>	<b>100.0%</b>	<b>2132</b>	<b>100.0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la quinta rama cuarta relación (tabla 15), se muestra asociación entre la víctima de robo edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado y la calificación del trabajo de la comisaria ( $X^2 = 29,24$ ;  $p < 0,001$ ;  $gl = 1$ ), con un odds ratio  $OR = 1,375$  (*I.C.* 95%: 1,225 – 1,543), significando que las

personas que califican como mala el trabajo de la comisaria presentan 1,375 veces de posibilidad de ser víctima de robo que las personas que califican como buena el trabajo de la comisaria. Así mismo, el 55,2% de las víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado califican como mala el trabajo de la comisaria, mientras que el 52,7% de las personas que no fueron víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado califican como buena el trabajo de la comisaria.

Tabla N° 16

*Víctima de robo con edad en alto riesgo de robo, forma segura de sentirse al caminar por la calle, que consideran que es seguro caminar por la zona o barrio sola y de noche, que van al cine o al teatro según hábito de haber dejado de llevar mucho dinero en efectivo*

Ha dejado de llevar mucho dinero en efectivo	Víctima de robo			
	No		Si	
	N	%	N	%
No / No lleva mucho dinero en efectivo	329	30.4%	187	40.3%
Si	754	69.6%	277	59.7%
<b>Total</b>	<b>1083</b>	<b>100.0%</b>	<b>464</b>	<b>100.0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la quinta rama cuarta relación (tabla 16), se muestra asociación entre la víctima de robo edad en alto riesgo de robo, forma segura de sentirse al caminar por la calle, que consideran que es seguro caminar por la zona o barrio sola y de noche, que van al cine o al teatro y el hábito de haber dejado de llevar mucho dinero en efectivo ( $X^2 = 14,39; p < 0,001; gl = 1$ ), con un odds ratio  $OR = 1,547$  ( $I.C. 95\%: 1,234 - 1,940$ ), significando que las personas que han cambiado sus hábitos de llevar mucho dinero en efectivo presentan 1,547 veces de posibilidad de ser víctima de robo que las personas que



no han cambiado sus hábitos de llevar mucho dinero en efectivo. Así mismo, el 40,3% de las víctimas de robo con edad en alto riesgo de robo, forma segura de sentirse al caminar por la calle, que consideran que es seguro caminar por la zona o barrio sola y de noche, que van al cine o al teatro han cambiado el hábito de llevar mucho dinero en efectivo, mientras que el 69,6% de las personas que no fueron víctimas de robo con edad en alto riesgo de robo, forma segura de sentirse al caminar por la calle, que consideran que es seguro caminar por la zona o barrio sola y de noche, que van al cine o al teatro no han cambiado el hábito de llevar mucho dinero en efectivo.

Tabla N° 17

*Víctima de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador /independiente /obrero /hogar, que no cambió su hábito y no lleva mucho dinero en efectivo, que no cambió si hábito de ir a los centros comerciales según forma de sentirse al caminar por la zona o barrio sola y de noche*

<b>Caminar por la zona o barrio sola y de noche</b>	<b>Víctima de robo</b>			
	<b>No</b>		<b>Si</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
Inseguro	2278	84,2%	1115	88,7%
Seguro	428	15,8%	142	11,3%
<b>Total</b>	<b>2706</b>	<b>100,0%</b>	<b>1257</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la sexta rama primera relación (tabla 17), se muestra asociación entre la víctima de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador /independiente /obrero /hogar, que no cambió su hábito y no lleva mucho dinero en efectivo, que no cambió si hábito de ir a los centros comerciales con la forma de sentirse al caminar por la zona o barrio sola y de noche ( $X^2 = 14,24$ ;  $p < 0,001$ ;  $gl = 1$ ), con un odds ratio  $OR = 1,475$  (*I. C.* 95%: 1,204 – 1,807), significando que las personas que se sienten inseguros de caminar por la zona o barrio sola y de noche presentan 1,4785 veces de posibilidad de ser víctima de robo que las

personas que se sienten seguros al caminar por la zona o barrio sola y de noche. Así mismo, el 88,7% de las víctimas de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador /independiente /obrero /hogar, que no cambió su hábito y no lleva mucho dinero en efectivo, que no cambió si hábito de ir a los centros comerciales se sienten inseguros al caminar por la zona o barrio sola y de noche, mientras que el 15,8% de las personas que no fueron víctimas de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador /independiente /obrero /hogar, que no cambió su hábito y no lleva mucho dinero en efectivo, que no cambió si hábito de ir a los centros comerciales se sienten seguros al caminar por la zona o barrio sola y de noche.

Tabla N° 18

*Víctima de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de trabajador no remunerado / No trabaja, que no cambió su hábito y no lleva mucho dinero en efectivo, que califica como mala el trabajo de la comisaria según cambio de hábito en dejar de ir al estadio*

Ha dejado de ir al estadio	Víctima de robo			
	No		Si	
	N	%	N	%
Si / No va al estadio	1143	73,3%	343	65,7%
No	416	26,7%	179	34,3%
<b>Total</b>	<b>1559</b>	<b>100,0%</b>	<b>522</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la sexta rama segunda relación (tabla 18), se muestra asociación entre la víctima de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de trabajador no remunerado / No trabaja, que no cambió su hábito y no lleva mucho dinero en efectivo, que califica como mala el trabajo de la comisaria y el cambio de hábito en dejar de ir al estadio ( $X^2 = 11,09; p = 0,003; gl = 1$ ), con un odds ratio  $OR = 1,434$  (I. C. 95%: 1,159 – 1,774), significando que las

personas que no cambiaron sus hábitos de ir al estadio presentan 1,434 veces de posibilidad de ser víctima de robo que las personas que si cambiaron sus hábitos de ir al estadio o simplemente no van. Así mismo, el 34,3% de las víctimas de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de trabajador no remunerado / No trabaja, que no cambió su hábito y no lleva mucho dinero en efectivo, que califica como mala el trabajo de la comisaria no han cambio el hábito en dejar de ir al estadio, mientras que el 73,3% de las personas que no fueron víctimas de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de trabajador no remunerado / No trabaja, que no cambió su hábito y no lleva mucho dinero en efectivo, que califica como mala el trabajo de la comisaria han cambiado de hábito en dejar de ir al estadio.

Tabla N° 19

*Víctima de robo con edad en bajo riesgo de robo, forma segura de sentirse al caminar por la calle, no cuentan con tarjetas de crédito o débito, sentirse seguros al caminar por la zona o barrio sola y de noche, han dejado y siguen tomando taxis en la calle según sexo de la víctima*

Sexo	Víctima de robo			
	No		Si	
	N	%	N	%
Hombre	489	50.5%	86	61.0%
Mujer	479	49.5%	55	39.0%
<b>Total</b>	<b>968</b>	<b>100.0%</b>	<b>141</b>	<b>100.0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la sexta rama tercera relación (tabla 19), se muestra asociación entre la víctima de robo con edad en bajo riesgo de robo, forma segura de sentirse al caminar por la calle, no cuentan con tarjetas de crédito o débito, sentirse seguros al caminar por la zona o barrio sola y de noche, han dejado y siguen tomando taxis en la calle y el sexo de la víctima ( $X^2 = 5,41; p = 0,020; gl = 1$ ), con un odds ratio  $OR = 1,532$  (I. C. 95%: 1,067 –

2,198), significando que las personas de sexo masculino presentan 1,532 veces de posibilidad de ser víctima de robo que las personas de sexo femenino. Así mismo, el 61,0% de las víctimas de robo con edad en bajo riesgo de robo, forma segura de sentirse al caminar por la calle, no cuentan con tarjetas de crédito o débito, sentirse seguros al caminar por la zona o barrio sola y de noche, han dejado y siguen tomando taxis en la calle son de sexo masculino, mientras que el 49,5% de las personas que no fueron víctimas de robo con edad en bajo riesgo de robo, forma segura de sentirse al caminar por la calle, no cuentan con tarjetas de crédito o débito, sentirse seguros al caminar por la zona o barrio sola y de noche, han dejado y siguen tomando taxis en la calle son de sexo femenino.

Tabla N° 20

*Víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo de sin nivel hasta secundaria, que no han dejado de frecuentar centros comerciales según forma de sentirse al caminar por la zona o barrio sola y de noche*

Caminar por la zona o barrio sola y de noche	Víctima de robo			
	No		Si	
	N	%	N	%
Inseguro	1980	83,5%	1327	88,4%
Seguro	390	16,5%	174	11,6%
<b>Total</b>	<b>2370</b>	<b>100,0%</b>	<b>1501</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la sexta rama cuarta relación (tabla 20), se muestra asociación entre la víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo de sin nivel hasta secundaria, que no han dejado de frecuentar centros comerciales y la forma de sentirse al caminar por la zona o barrio sola y de noche ( $X^2 = 17,46; p < 0,001; gl = 1$ ), con un odds ratio  $OR = 1,502$  (I.C. 95%: 1,240 – 1,820), significando que las personas que se sienten inseguras al caminar por la zona o barrio sola y de noche presentan

1,502 veces de posibilidad de ser víctima de robo que las personas que se sienten seguros al caminar por la zona o barrio sola y de noche. Así mismo, el 88,4% de las víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo de sin nivel hasta secundaria, que no han dejado de frecuentar centros comerciales se sienten inseguros al caminar por la zona o barrio sola y de noche, mientras que el 16,5% de las personas que no fueron víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo de sin nivel hasta secundaria, que no han dejado de frecuentar centros comerciales se sienten seguros al caminar por la zona o barrio sola y de noche.

Tabla N° 21

*Víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado, con mala calificación del trabajo de la comisaria según sexo de la víctima*

Sexo	Víctima de robo			
	No		Si	
	N	%	N	%
Hombre	584	48,5%	492	41,8%
Mujer	619	51,5%	685	58,2%
<b>Total</b>	<b>1203</b>	<b>100,0%</b>	<b>1177</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la sexta rama quinta relación (tabla 21), se muestra asociación entre la víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado, con mala calificación del trabajo de la comisaria y el sexo de la víctima ( $X^2 = 10,92; p = 0,001; gl = 1$ ), con un odds ratio  $OR = 1,314$  (I. C. 95%: 1,117 –

1,544), significando que las personas de sexo femenino presentan 1,314 veces de posibilidad de ser víctima de robo que las personas de sexo masculino. Así mismo, el 58,2% de las víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado, con mala calificación del trabajo de la comisaria son sexo femenino, mientras que el 48,5% de las personas que no fueron víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, que no llevan y no han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado, con mala calificación del trabajo de la comisaria son de sexo masculino.

Tabla N° 22

*Víctima de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador / independiente / obrero / hogar, no llevan y no han dejado de llevar mucho dinero en efectivo, que no han dejado de frecuentar centros comerciales, con forma insegura de sentirse al caminar por la zona o barrio sola y de noche según existencia de vigilancia de la Policía Nacional del Perú*

Existe vigilancia de la Policía Nacional del Perú	Víctima de robo			
	No		Si	
	N	%	N	%
Si / No hay PNP	1035	45,4%	442	39,6%
No	1243	54,6%	673	60,4%
<b>Total</b>	<b>2278</b>	<b>100,0%</b>	<b>1115</b>	<b>100,0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la séptima rama primera relación (tabla 22), se muestra asociación entre la víctima de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador / independiente / obrero / hogar, no llevan y no han dejado de llevar mucho dinero en efectivo, que no han dejado de frecuentar centros comerciales, con forma insegura de sentirse al caminar por la zona o barrio sola y de noche y la existencia de vigilancia de la Policía Nacional del Perú ( $X^2 = 10,22; p =$

0,004;  $gl = 1$  ), con un odds ratio  $OR = 1,268$  (I.C. 95%: 1,096 – 1,467) , significando que las personas que consideran que no existe vigilancia de la PNP presentan 1,268 veces de posibilidad de ser víctima de robo que las personas que consideran que si existe vigilancia de la PNP. Así mismo, el 60,4% de las víctimas de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador / independiente / obrero / hogar, no llevan y no han dejado de llevar mucho dinero en efectivo, que no han dejado de frecuentar centros comerciales, con forma insegura de sentirse al caminar por la zona o barrio sola y de noche consideran que no existe vigilancia de la Policía Nacional del Perú, mientras que el 45,4% de las personas que no fueron víctimas de robo con edad en bajo riesgo de robo, forma insegura de sentirse al caminar por la calle, con ocupación principal de empleador / independiente / obrero / hogar, no llevan y no han dejado de llevar mucho dinero en efectivo, que no han dejado de frecuentar centros comerciales, con forma insegura de sentirse al caminar por la zona o barrio sola y de noche consideran que existe vigilancia de la Policía Nacional del Perú.

Tabla N° 23

*Víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, no lleva y no ha dejado de llevar mucho dinero en efectivo, con nivel educativo de sin nivel hasta secundaria, que no ha dejado de frecuentar centros comerciales, sentirse inseguro al caminar por la zona o barrio sola y de noche según el hábito de salir de noche*

Ha dejado de salir de noche	Víctima de robo			
	No		Si	
	N	%	N	%
Si	337	17.0%	290	21.9%
No / No sale de noche	1643	83.0%	1037	78.1%
<b>Total</b>	<b>1980</b>	<b>100.0%</b>	<b>1327</b>	<b>100.0%</b>

**Fuente:** Base de datos de la Encuesta Nacional Especializada sobre Victimización 2017

En la séptima rama segunda relación (tabla 23), se muestra asociación entre la víctima de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, no lleva y no ha dejado de llevar mucho dinero en efectivo, con nivel educativo de sin nivel hasta secundaria, que no ha dejado de frecuentar centros comerciales, sentirse inseguro al caminar por la zona o barrio sola y de noche con el hábito de salir de noche ( $X^2 = 12,08; p = 0,002; gl = 1$ ), con un odds ratio  $OR = 1,363$  (I. C. 95%: 1,144 – 1,624), significando que las personas que cambiaron su hábito de salir por las noches presentan 1,363 veces de posibilidad de ser víctima de robo que las personas que no salen o no cambiaron sus hábitos de salir por las noches.

Así mismo, de la tabla 20 se observa que el 21,9% de las víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, no lleva y no ha dejado de llevar mucho dinero en efectivo, con nivel educativo de sin nivel hasta secundaria, que no ha dejado de frecuentar centros comerciales, sentirse inseguro al caminar por la zona o barrio sola y de noche han cambiado el hábito de salir de noche, mientras que el 83,0% de las personas que no fueron víctimas de robo con edad en alto riesgo de robo, forma insegura de sentirse al caminar por la calle, no lleva y no ha dejado de llevar mucho dinero en efectivo, con nivel educativo de sin nivel hasta secundaria, que no ha dejado de frecuentar centros comerciales, sentirse inseguro al caminar por la zona o barrio sola y de noche, no salen o no han cambiado el hábito de salir de noche.

## **EVALUACION**

Para el análisis del algoritmo CHAID trabajó se trabajó con el 100% de los registros (37820 registros), ingresando 31 variables al análisis, de los cuales 17 variables fueron significativas, siendo las más importante la edad de la víctima.



## **IMPLANTACION**

Respecto al análisis del algoritmo CHAID, se observa que las víctimas de robo presentan perfiles significativos las cuales deben servir de base para concientizar a las personas a tener mayor cuidado y no ser objetivos claros para los delincuentes.

### **5.2. DISCUSION DE RESULTADOS**

El Algoritmo CHAID es una herramienta que no requiere conocer su distribución de datos de las variables estudiadas (Rayo & Cortes, 2010) lo cual facilita su utilización para búsqueda, identificación y obtención de las relaciones existentes entre los predictores de las víctimas de robo del Perú existentes en una gran cantidad de datos almacenados de la Encuesta Nacional Especializada sobre Victimización 2017 (Torres, C., 2015).

De las 31 variables identificadas como predictores de las víctimas de robo del Perú, sólo fueron significativas 17 variables, probablemente debido a que el Algoritmo CHAID presenta una limitación muy importante que puede no capturar todas las interacciones entre las variables predictores y la variable objetivo (Moreno, A.; Vicente, P. & Galindo, P., 2016) Sin embargo, estas 17 variables predictores significativas representan las características de las víctimas de robo en el Perú, con OR (Odds Ratio) superiores a la unidad, representando así que son variables predictores de riesgo. Estas variables predictoras demuestran la existencia de dependencia con la víctima de robo en el Perú (Escobar, M., 1998), siendo el predictor de mayor importancia la edad de la persona. La edad es una característica principal para ser víctima de robo en Perú, la cual está sujeta a que si la persona es muy joven o muy viejo tiene menor posibilidad de ser elegido como víctima (OR = 1,890), lo cual no coincide con los encontrado por Cárdenas, O. (2012) donde la edad es un atributo personal no correlacionado con la probabilidad de ser

víctima. Ello puede deberse al contexto social y cultural en los que se realizaron los estudios.

Por otro lado, la forma de sentirse al caminar por la calle representa un predictor de las víctimas de robo, ya que al sentirse inseguro presenta mayor posibilidad de ser víctima de robo en Perú (OR = 1,369), lo cual coincide con Quinteros, D. (2014) donde encontró que la inseguridad tiene mayor chace para ser víctima de robo por la carencia de esquemas de protección (El Instituto Nacional de Estadística y Geografía de México, 2015).

El caminar por la zona o barrio sola y de noche es una característica predictora muy probable de ser víctima de robo en Perú, por la inseguridad (Quinteros, D., 2014) y porque no se cuentan con mecanismos de protección para la persona en ese horario (El Instituto Nacional de Estadística y Geografía de México, 2015), estas características favorecen la exposición al riesgo de ser víctima de robo.

El hábito de llevar mucho dinero en efectivo es otro de los predictores importantes para ser víctima de robo en Perú, ya que la característica económica dinamiza la exposición a ser víctima de robo (Ministerio de Defensa Nacional – Policía Nacional, 2016), ello sumado a las pocas previsiones de las personas al movilizarse con efectivo.

Finalmente, el quinto predictor de mayor importancia es el nivel educativo de las personas, cabe resaltar que aquellas personas con estudios inferiores a la superior son poco llamativos y atractivos para ser víctimas de robo, mientras que las personas con nivel superior sean técnico o universitario e incluido el postgrado son atractivos para ser víctimas de robo, dado que sus hábitos de vestimenta son diferenciados en el día a día.

El perfil más riesgoso de la víctima de robo en Perú está caracterizado por: presentar una edad alta de riesgo de robo (menor a 18 años de edad y mayores a 40 años de edad), sentirse inseguro al caminar por la calle que, si han dejado de llevar mucho dinero en

efectivo, con nivel educativo superior a postgrado y mala calificación del trabajo de la comisaria.

El segundo perfil de riesgo para ser víctima de robo es sentirse inseguro al caminar por la calle, tener como ocupación principal de empleado u otra ocupación y ser de sexo femenino.

El tercer perfil de riesgo para ser víctima de robo es la inseguridad de caminar por la zona o barrio sola y de noche y tener un nivel educativo superior a postgrado, así como, llevar consigo tarjetas de crédito o débito como un hábito y una mala calificación del trabajo de la comisaria.

El cuarto perfil de riesgo para ser víctima de robo es tener un nivel educativo de superior a postgrado y una mala calificación del trabajo de la comisaria, tener hábitos de ir al cine o teatro y haber cambiado su hábito de llevar mucho dinero en efectivo, inseguridad al caminar por la zona o barrio sola y de noche y contar con un empleo o trabajo.

El quinto perfil de riesgo para ser víctima de robo es la inseguridad al caminar por la zona o barrio sola y de noche y la no existencia de vigilancia de la Policía Nacional del Perú, así como el hábito de salir por la noche.

Finalmente, con la aplicación del Algoritmo CHAID se pone en manifiesto dichos patrones de comportamiento a nivel nacional, aplicando a priori toda la metodología de la minería de datos, los cuales reflejas la realidad estudiada (Ramos, S., 2015; Torres, C., 2015 y Rayo & Cortes, 2010).

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

## CONCLUSIONES

- Se identificaron 17 variables predictores de las víctimas de robo del Perú mediante el algoritmo CHAID, según la Encuesta Nacional Especializada sobre Victimización 2017, las cuales son: la edad, cómo se siente caminando por la calle, caminar por la zona o barrio sola y de noche, ha dejado de llevar mucho dinero en efectivo, nivel educativo alcanzado, calificación del trabajo de la comisaria, ha dejado de frecuentar centros comerciales, ha dejado de ir al cine o al teatro, el sexo, ocupación principal o negocio, ha dejado de ir a la Institución Educativa, Institutos Tecnológicos, Universidad, etc., ha dejado de llevar consigo tarjetas de crédito o débito, ha dejado de salir de noche, ha dejado de tomar taxis en la calle, ha dejado de ir al estadio, condición laboral y la existencia de vigilancia de la Policía Nacional del Perú.
- Las relaciones bivariadas entre las variables predictoras de las víctimas de robo del Perú son significativas inclusive al 95% de confiabilidad, de acuerdo a la prueba Chi-cuadrado de Pearson.
- Con la aplicación del Algoritmo CHAID a los predictores de las víctimas de robo del Perú, según la Encuesta Nacional Especializada sobre Victimización 2017 se encontró el perfil de la víctima de robo en Perú más riesgosa y está caracterizado por: presentar una edad alta de riesgo de robo (menor a 18 años de edad y mayores a 40 años de edad), sentirse inseguro al caminar por la calle, que si han dejado de llevar mucho dinero en efectivo, con nivel educativo superior a postgrado y mala calificación del trabajo de la comisaria.

## 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 y algoritmo CHAID.

## **RECOMENDACIONES**

A la ciudadanía para que tome en cuenta respecto a las características de mayor riesgo al robo, y tomen las precauciones del caso a fin de no ser presas fáciles de los victimarios.

A los estudiantes de la Escuela Académico Profesional de Estadística e Informática, para que profundicen sus conocimientos en temas de minería de datos y algoritmo CHAID.

## REFERENCIAS BIBLIOGRAFICAS

- AmeticaTV (27.09.2018). *INEI: 26 de cada 100 peruanos fueron víctimas de robos entre marzo y agosto de 2018*. Descargado de: <https://www.americatv.com.pe/noticias/actualidad/inei-26-cada-100-peruanos-fueron-victimas-hechos-delictivos-entre-marzo-y-agosto-2018-n340565>
- Bouroche, J. M. y Tennenhaus, M. (1972). *Some Segmentation Methods*, Metra, 7, pp. 407-418.
- Cárdenas Rodríguez, Oscar Javier. (2012). *¿Con qué probabilidad me toca ser víctima de un delito?*. EconoQuantum, 9(1), 171-187. Recuperado en 03 de octubre de 2018, de [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S1870-66222012000100011&lng=es&tlng=es](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-66222012000100011&lng=es&tlng=es).
- Chaux, E.; Camargo, D. C.; León, M. & Trujillo, D. (2013). *Actitudes y dinámicas de robo en un colegio de nivel socioeconómico medio-alto*. *Revista Criminalidad*, enero-abril, Vol. 55 (1), pp. 11-29.
- Cohen Lawrence y Felson Marcus (1979). *Social change and crime rate trends: a routine activity approach*. *American Sociological Review*, Vol. 44, (Agosto 1979): 588-608.
- Díaz-Pérez & Bethencourt-Cejas (2016). *CHAID algorithm as an appropriate analytical method for tourism market segmentation*. *Journal of destination marketing & management*. 5.pp.275-282.
- El Instituto Nacional de Estadística y Geografía de México (2015). *En Números, Documentos de Análisis y Estadísticas*. México.
- El Ministerio de Defensa Nacional – Policía Nacional (2016). *Hurto a personas y su afectación en la economía y contexto social*. Colombia.



- Escobar, M. (1998). *Las aplicaciones del análisis de segmentación: el procedimiento CHAID*. Working Paper número 31 publicado por el Centro de Estudios Avanzados en Ciencias Sociales del Instituto Juan March de Estudios e Invesigaciones de enero de 1992.
- Esquerda, A. y Trujillano, J. (2010). *Árboles de clasificación para la generación de reglas de decisión clínica*. Descargado de: [www.jano.es](http://www.jano.es)
- Franco, A.; Carrasco, J.; Sánchez, G. & Martínez, J. (2008). *Árboles de Decisión para Grandes Conjuntos de Datos*. Reporte Técnico No. CCC-08-001. Coordinación de Ciencias Computacionales INAOE. Mexico.
- Galindo, M.; Vicente, J.; Dorado, A.; Vicente, P. y Patino, M. (2009). *Una alternativa al algoritmo chaid de segmentación basada en entropía*. Revista de Matemática: Teoría y Aplicaciones 2010 17(2) : 179–197
- Gomes Fabio y Lourenco S. Paz (2008). *The Determinants of Criminal Victimization in Sao Paulo state-Brazil*. Brazilian Review of Econometrics, Núm 28, Vol. 2 217-238.
- Hunter Albert (1978). *Symbols of incivility: social disorder and fear of crime in urban neighborhoods*. Reactions to Crime Project. Washington, D.C: U.S. Department of Justice National Criminal Justice Reference Service
- IBM Knowledge Center (2018). <https://www.ibm.com/support/knowledgecenter/es>
- Kass, G. V. (1980). *An Exploratory Technique for Investigating Large Quantities of Categorical Data*, Applied Statistics, 29, pp. 119-127
- Madgison, J. (1993): *SPSS for Windows CHAID reléase 6.0*, Chicago, SPSS.
- Meier Robert y Miethe Terance (1993). *Understanding Thories of Criminal Victimization*. Crime and Justice, Vol. 17: 459-499

- Moreno, A.; Vicente, P. & Galindo, P. (2016). *Aprendizaje basado en árboles de decisión: un estudio crítico desde Weka, RapidMiner y SPSS Modeler*. XXVI Simposio Internacional de Estadística 2016, Sincelejo, Sucre, Colombia, 8 al 12 de Agosto de 2016.
- Morgan, J. N. y Sonquist, J. A. (1963): *Problems in the Analysis of Survey Data*, Journal of the American Statistical Association, 58, pp. 415-4
- Navarro, J. & Pastor, E. (2017). *Factores dinámicos en el comportamiento de delincuentes juveniles con perfil de ajuste social. Un estudio de reincidencia*. Psychosocial intervention, 26:19-27.
- Outlaw Maureen, Ruback Barry y Britt Chester (2002). *Repeat and Multiple Victimizations: Th Role of Individual and Contextual Factors*. Estados Unidos: U.S. Department of Justice.
- Quinteros, D. (2014). *Delitos del espacio público y el problema de la “cifra negra”*: una aproximación a la no-denuncia en Chile. Polít. crim. Vol. 9, N° 18 (Diciembre 2014), Art. 1 2, pp. 691-712.
- Ramos, S. (2015). *TAID versus CHAID Búsqueda de perfiles de mujeres trabajadoras en el servicio doméstico*. Tesis para optar el grado de maestro en Análisis Avanzado de Datos Multivariantes de la Universidad de Salamanca – España.
- Rayo & Cortes (2010). *Aplicación del Chaid para identificar las características económico financieras de las empresas inmobiliarias más rentables en España*. Journal of Economics, Finance and Administrative Science, 15(29).
- Sánchez, O. & Fuentes, C. (2015). *El robo de vehículos y su relación espacial con el contexto sociodemográfico en tres delegaciones centrales de la Ciudad de México (2010)*. Investigaciones geográficas, boletín N° 89, Instituto de Geografía, UNAM, pp. 107-120.

- Sanz E. y Ponce de León, A. (2010). *Claves en la aplicación del algoritmo chaid. Un estudio del ocio físico deportivo universitario*. Revista de Psicología del Deporte. 2010. Vol. 19, núm. 2, pp. 319-333.
- Torres, C. (2015). *Métodos de investigación del mercado de créditos de consumo, mediciones de los patrones determinantes según su destino aplicado a instituciones financieras*. Tesis para optar el grado de Maestro en Administración de Empresas de la Universidad del Pacífico, Guayaquil – Colombia.
- Trujillano, J.; Sarria, A.; Esquerda, A.; Badia, M.; Palma, M. y March, J. (2008). *Aproximación a la metodología basada en árboles de decisión (CART). Mortalidad hospitalaria del infarto agudo de miocardio*. Gac Sanit, 22(1):65-72.
- Vivanco, M. (1999). *Análisis estadístico multivariable: teoría y práctica*. Editorial Universitaria S.A., Primera Edición. Chile.

## **ANEXOS**



# ENCUESTA NACIONAL ESPECIALIZADA SOBRE VICTIMIZACIÓN 2017



INFORMACIÓN CONFIDENCIAL AMPLIADA POR DECRETO SUPLENTE 043 - 2001 - PCM DEL SECRETO DE ESTADO

<b>DOC.ENEVIC.01</b>	
CUESTIONARIO N°	Cuestionario Adicional 1

CONGLOMERADO N°	N° DE SELECCIÓN DE LA VIVIENDA
-----------------	--------------------------------

¿ES UNA VIVIENDA DE REEMPLAZO?	
SI..... 1	→ Anote el número de selección de la vivienda reemplazada
No..... 2	

<b>UBICACIÓN GEOGRÁFICA</b>	
1. DEPARTAMENTO	
2. PROVINCIA	
3. DISTRITO	
4. CENTRO POBLADO	

<b>UBICACIÓN MUESTRAL</b>	
5. ZONA N°	
6. MANZANA N°	
7. VIVIENDA N°	

VISITA	ENCUESTADOR/A				RESULTADO DE LA VISITA (*)
	FECHA	HORA DE	A	HORA	
Primera					
segunda					
Tercera					
Cuarta					
Quinta					
Sexta					

8. DIRECCIÓN DE LA VIVIENDA

Tipo de vía: Avenida ..... 1 Calle ..... 2 Jirón ..... 3 Pasejo ..... 4 Carretera ..... 5 Otro ..... 6

Nombre de la Calle, Av., Jr., Carretera, etc.	N° DE PUERTA	BLOCK	INT.	PISO	MZ.	LOTE	KIL.	TELÉFONO
---	--------------	-------	------	------	-----	------	------	----------

9. TOTAL DE HOGARES QUE OCUPAN LA VIVIENDA

INICIE LA ENTREVISTA CON EL/LA JEFE/A DEL HOGAR	
Señoría: Si HOGAR es la persona o grupo de personas que se alimentan de una misma olla y atienden en común otras necesidades básicas.	¿Cuántos hogares ocupan esta vivienda?

10. N° DE HOGAR
-----------------

11. NOMBRES Y APELLIDOS DE EL/LA JEFE/A DEL HOGAR

RESPONSABLE	DNI	NOMBRES Y APELLIDOS
ENCUESTADOR/A		
SUPERVISOR/A LOCAL		
COORDINADOR/A DEPARTAMENTAL		
SUPERVISOR/A NACIONAL		

<b>13. RESULTADO FINAL DE LA ENCUESTA</b>	
FECHA	
RESULTADO	

<b>(*) CÓDIGO 8 DE RESULTADO</b>	
1. COMPLETA	5. VIVIENDA DESOCUPADA
2. INCOMPLETA	6. NO SE INICIO LA ENTREVISTA
3. RECHAZO	7. OTRO
4. AUSENTE	(Especifique)

<b>15. IDIOMA DE LA ENTREVISTA</b> (Círcule uno o más códigos)	
Castellano .....	1
Quechua .....	2
Otro .....	3
(Especifique)	

<b>16. RESULTADO DE LA VICTIMIZACIÓN</b> (Acontecidos en los últimos 12 meses)	
¿Vivienda con víctimas?	Si No
1	2



**FORMATO DE AUTORIZACIÓN PARA PUBLICACIÓN DE TESIS Y TRABAJOS DE INVESTIGACIÓN,  
PARA OPTAR GRADOS ACADÉMICOS Y TÍTULOS PROFESIONALES EN EL  
REPOSITORIO INSTITUCIONAL DIGITAL - UNASAM**

Conforme al Reglamento del Repositorio Nacional de Trabajos de Investigación - RENATI,  
Resolución del Consejo Directivo de SUNEDU N° 033-2016-SUNEDU/CD

**1. Datos del Autor:**

Apellidos y Nombres: Alvarado Lazaro Wang Marco

Código de alumno: 081.0403.288

Teléfono: 927202562

Correo electrónico: [marcoalvarado321@gmail.com](mailto:marcoalvarado321@gmail.com) DNI o Extranjería: 46995866

**2. Modalidad de trabajo de investigación:**

Trabajo de investigación

Trabajo académico

Trabajo de suficiencia profesional

Tesis

**3. Título profesional o grado académico:**

Bachiller

Título

Segunda especialidad

Licenciado

Magister

Doctor

**4. Título del trabajo de investigación:**

Algoritmo CHAID aplicado a los predictores de las víctimas de robo del Perú, según la encuesta Nacional Especializada sobre Victimización 2017.

**5. Facultad de: Ciencias**

**6. Escuela, Carrera o Programa: Estadística e Informática.**

**7. Asesor:**

Apellidos y Nombres: Mendoza López Angel Deciderio

Teléfono: 943225752

Correo electrónico: [amacho@yahoo.com](mailto:amacho@yahoo.com)

DNI o Extranjería: 17824554

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 impresa y digital, son las versiones finales del trabajo sustentado y aprobado por el jurado y son de autoría del suscrito en estricto respeto de la legislación en materia de propiedad intelectual.

Firma: .....

D.N.I.: 46995866

FECHA: 12 / AGO / 2019



**FORMATO DE AUTORIZACIÓN PARA PUBLICACIÓN DE TESIS Y TRABAJOS DE INVESTIGACIÓN,  
PARA OPTAR GRADOS ACADÉMICOS Y TÍTULOS PROFESIONALES EN EL  
REPOSITORIO INSTITUCIONAL DIGITAL - UNASAM**

Conforme al Reglamento del Repositorio Nacional de Trabajos de Investigación - RENATI,  
Resolución del Consejo Directivo de SUNEDU N° 013-2016-SUNEDU/CJ

**1. Datos del Autor:**

Apellidos y Nombres: ZAVALA SALVADOR Darwin James

Código de alumno: 081.0403.301                      Teléfono: 986736472

Correo electrónico: [djzavala51@gmail.com](mailto:djzavala51@gmail.com)                      DNI o Extranjería: 47001026

**2. Modalidad de trabajo de investigación:**

- Trabajo de investigación                       Trabajo académico  
 Trabajo de suficiencia profesional                       Tesis

**3. Título profesional o grado académico:**

- Bachiller                       Título                       Segunda especialidad  
 Licenciado                       Magister                       Doctor

**4. Título del trabajo de investigación:**

Algoritmo CHAID aplicado a los predictores de las víctimas de robo del Perú, según la encuesta Nacional Especializada sobre Victimización 2017.

**5. Facultad de:** Ciencias

**6. Escuela, Carrera o Programa:** Estadística e Informática.

**7. Asesor:**

Apellidos y Nombres: Mendoza López Angel Decideno                      Teléfono: 943225752

Correo electrónico: [amendoza@ralton.com](mailto:amendoza@ralton.com)                      DNI o Extranjería: 17824554

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 impresa y digital, son las versiones finales del trabajo sustentado y aprobado por el jurado y son de autoría del suscrito en estricto respeto de la legislación en materia de propiedad intelectual.

Firma: 

D.N.I.: 47001026

FECHA: 12 / AGO / 2019





UNIVERSIDAD NACIONAL  
SANTIAGO ANTÚNEZ DE MAYOLO



"Una Nueva Universidad para el Desarrollo"  
ESCUELA PROFESIONAL DE ESTADÍSTICA E INFORMÁTICA  
AV. CENTENARIO N° 200 - TELÉFONO (043) 649020 ANEXO 1913  
HUARAZ - ANCASH - PERÚ

*Año de la lucha contra la corrupción y la impunidad*

## ACTA DE SUSTENTACIÓN DE TESIS

Los Miembros del Jurado de la Revisión y Sustentación de Tesis de la Escuela Académico Profesional de Estadística e Informática de la Facultad de Ciencias, designados mediante Resolución de Consejo de Facultad N° 0293-2018-UNASAM-FC, se reunieron el día martes 02 de julio de 2019, a horas 4:30 p.m. en el Aula G° 301 de la Facultad de Ciencias en acto público para evaluar la Sustentación de Tesis, presentado por los:

➤ Bachilleres :

- DARWIN JAMES ZAVALA SALVADOR
- WANG MARCO ALVARADO LÁZARO

Tesis Titulada : "ALGORITMO CHAID APLICADO A LOS PREDICTORES DE LAS VICTIMAS DE ROBO DEL PERÚ, SEGÚN LA ENCUESTA NACIONAL ESPECIALIZADA SOBRE VICTIMIZACIÓN 2017",

Después de la Sustentación y las respuestas a las preguntas, el Jurado lo declara APROBADO para optar el Título Profesional de Licenciado en Estadística e Informática, con el calificativo de Diecisiete (17)

En señal de conformidad y para constancia, firmamos la presente ACTA, siendo las 5:30 pm del mismo día y año.

Huaraz, 02 de julio de 2019.

MSc. JORGE LUIS LLANOS TIZNADO  
PRESIDENTE

Dr. ROGER PEDRO NORABUENA FIGUEROA  
SECRETARIO

MSc. HUGO WALTER MALDONADO LEYVA  
VOCAL

DR. ANGEL DECIDERIO MENDOZA LOPEZ  
ASESOR

