

Memorias



Sociedad Latinoamericana en
Percepción Remota y Sistemas
de Información Espacial
Capítulo Colombia

30
AÑOS

Medellín, Colombia
29 de Septiembre al 3 de Octubre de 2014



**Clasificación de una Nube de Puntos LIDAR a Partir de Técnicas de Minería de Datos
Mediante el Uso de Software Libre**

**Classification of LIDAR Point Cloud Starting from Data Mining Techniques Using
Open Source Software**

Indira Paola Pachón Cendales #^{*1}

Salomón Einstein Ramírez Fernández#^{*2}

*Maestría en Ciencias de la Información y las Comunicaciones Énfasis en Geomática,
Universidad Distrital Francisco José de Caldas,
Carrera 7 No. 40B - 53 Bogotá D.C, Colombia*

* *Instituto Geográfico Agustín Codazzi – IGAC, Centro de Investigación y Desarrollo en
Información Geográfica – CIAF, Grupo de Percepción Remota y Aplicaciones*

Geográficas, Carrera 30 No. 48 – 51 Bogotá D.C, Colombia

¹ indira.pachon@igac.gov.co , ² salomon.ramirez@igac.gov.co

Palabras Clave: Clasificación, Tecnología LIDAR, Árboles de Decisión, Software Libre.

Resumen: El presente artículo muestra los resultados de la aplicación de técnicas propias de la minería de datos en la clasificación o filtrado de datos que contienen información del terreno proveniente de una nube de datos LIDAR (Light Detection and Ranking), de la cual se realizó la extracción de 7 clases informacionales reconocidas sobre el área de estudio (Sin clasificar, suelo, vegetación baja, vegetación media, vegetación alta, edificios y vías), estas clases fueron generalizadas de una gran variedad de clasificación de cobertura del suelo encontrada dentro de la nube de datos en consecuencia con la clasificación expuesta por la Sociedad Americana de Fotogrametría y Percepción Remota/Teledetección ASPRS,

y cuyo procesamiento fue abordado mediante la metodología para proyectos de minería de datos KDD (Knowledge Discovery Database).

La metodología comprende los siguientes pasos: Primero la eliminación de ruido del conjunto de datos, seguido de la obtención de una muestra tanto para entrenamiento como para validación mediante filtrado manual, y el filtrado automático de la nube de puntos mediante clasificación supervisada empleando el algoritmo Árboles de Decisión, y finalmente la evaluación de la exactitud del filtrado. Los resultados derivados sugieren una buena exactitud en el filtrado obtenido como resultado de la aplicación de esta técnica.

Key Words: Classification, LIDAR Technology, Decision Tree, Open Source.

Abstract: This paper presents the results of the application's own data mining techniques in sorting or filtering data containing terrain information from a cloud of LIDAR data (Light Detection and Ranking), which was performed extraction 7 recognized informational classes on the study area (Unclassified soil, low vegetation, medium vegetation, high vegetation, buildings and roads), these classes were widespread in a variety of classification of land cover found within the cloud data accordingly to the classification set forth by the American Society for Photogrammetry and Remote Sensing / Remote Sensing ASPRS, whose processing was addressed by the methodology for mining data from KDD (Knowledge Discovery Database).

The methodology includes the following steps: First removing noise in the dataset, followed by obtaining a sample for both training and validation by manual filtering, and automatic filtering the point cloud by using supervised classification algorithm Trees

decision, and finally the evaluation of the accuracy of the filter. The results suggest good accuracy resulting in the filtrate obtained as a result of the application of this technique.

1. Introducción

La tecnología LIDAR (Light Detection and Ranking) se puede definir como un sistema de teledetección activo de captura, que procesado de manera adecuada puede proporcionar información muy útil en diferentes áreas del conocimiento (Lanig & Zipf, 2009).

Dentro de las múltiples aplicaciones en las que se puede emplear esta tecnología, se encuentran desde aplicaciones de uso militar, topografía y aplicaciones medioambientales, las cuales son objeto del presente estudio. Las ventajas de esta tecnología frente a la producción cartográfica convencional, incide principalmente en la disminución de los costos de producción y en un aumento de la precisión de los puntos obtenidos.

De acuerdo con esto, el aumento en el empleo de esta tecnología es realmente significativo (Lanig & Zipf, 2009), por lo que cada vez más aparecen en el mercado gran variedad de software privativo y libre que permiten realizar el proceso y filtrado de esta información, aunque pese a sus múltiples ventajas, requiere de un tratamiento especial y un pre proceso de la información (Lanig et al. 2008), lo cual involucra un elevado costo computacional para obtener los resultados requeridos, sin dejar de lado que muchos de estos software son costosos y de uso complejo.

Por tal razón, en el presente estudio de dominio de aplicación se propone la implementación de técnicas de minería de datos para la clasificación o filtrado de la información del terreno contenida en una nube de datos LIDAR, y para lo cual se propone la siguiente hipótesis, “Es posible realizar filtrado de las diferentes capas de información

del terreno (Suelo, vegetación baja, vegetación media, vegetación alta, edificaciones y vías) de una nube de datos LIDAR mediante el uso de técnicas de minería de datos bajo una plataforma de software libre que garantice un nivel de confiabilidad de filtrado y clasificación, rendimientos computacionales y bajo costo del proceso”.

El presente estudio se plantea bajo las fases metodológicas que emplea la metodología para proyectos de minería de datos KDD como se muestra en la *Figura 1* (Ocaranza & Quiroz, 2005), en donde se plantea inicialmente el objeto de estudio, el cual se mencionó anteriormente, en una segunda fase se determina y describe el conjunto de datos objetivo, el cual corresponde a una nube LIDAR localizada en el municipio de San Alberto en el departamento del Cesar, ubicado al norte de Colombia, la tercera fase comprende la limpieza de datos y pre procesamiento de la información por medio del uso del *Software Libre R*, seguida de la fase de reducción y limpieza de datos, en la quinta fase de esta metodología se propone y justifica la técnica de minería de datos a usar para la explotación de la información, la cual corresponde a las técnicas de árboles de decisión donde se aplicó el algoritmo contenido en la librería *rpart de R* la cual se basa en el algoritmo de árboles bayesianos (Therneau et al. 2014), y finalmente se realiza la interpretación y análisis de los resultados y se observa si a partir de las técnicas usadas se cumplió con los objetivos de explotación propuestos inicialmente.

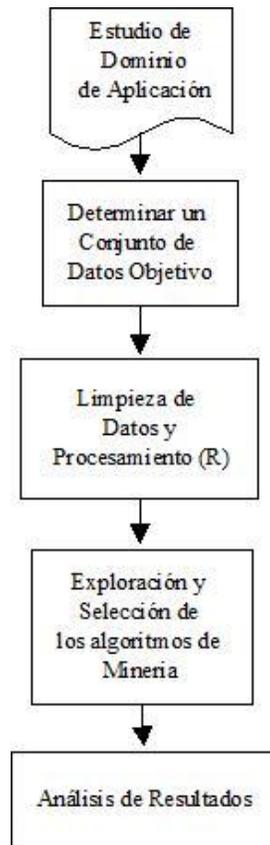


Figura 1. Metodología para proyectos de minería datos KDD

1.1. Trabajos Relacionados

La investigación de trabajos relacionados se realizó con el fin de conocer si se han usado este tipo de técnicas dentro del mismo contexto que se encuentra el presente estudio, partiendo de ello se encontraron varios autores los cuales proponen la clasificación de cobertura terrestre a partir de nubes de datos LIDAR aplicando las técnicas de minería de datos pero mediante el uso de algoritmos diferentes al expuesto para el estudio.

La propuesta metodológica planteada por Chehata et al. (2009) & Niemeyer et al. (2013) en donde proponen la clasificación de una nube de puntos LIDAR con el propósito de

integrar un clasificador de bosques aleatorios en un campo condicional (CRF) en el uso de la clasificación de múltiples clases, CRF ha demostrado que entrega buenos resultados en este tipo de clasificaciones. Aunque los autores proponen el uso de este algoritmo para la clasificación de complejas escenas urbanas donde se debe realizar la clasificación de múltiples clases (Chehata et al. 2009; Niemeyer et al. 2013) .

Zhang et al. (2013) proponen un método de clasificación basado en objetos para la clasificación de las nubes de puntos LIDAR en las zonas urbanas. En el proceso de clasificación, emplearon el algoritmo de agrupamiento para hacer que la remoción de valores atípicos de las nubes de puntos (Zhang et al. 2013). Trece características de la geometría, radiometría, la topología y las características de eco se calcularon, e hicieron el uso de máquinas de soporte vectorial (SVM) para clasificar los segmentos, y se propone el análisis del componente 3D para optimizar los resultados de la clasificación original. De acuerdo a las conclusiones obtenidas los autores sugieren que el método es capaz de hacer una clasificación de las nubes de puntos urbanas con la precisión global de clasificación mayor que 92,34 % y el coeficiente Kappa mayor que 0,8638, y la precisión de la clasificación se promueve con el aumento de la densidad de punto, que es significativo para varios tipos de aplicaciones (Zhang et al. 2013). A pesar que en este estudio se usó otro tipo de algoritmo se llegan a resultados y conclusiones muy similares en cuanto a la clasificación.

Otra propuesta realizada recientemente es la expuesta por los autores Collin et al. (2011), estos autores proponen el empleo de técnicas de aprendizaje automatizado en la clasificación de ecosistemas marinos, algoritmos de clasificación como Naive Bayes, Árboles de decisión, C4.5, bosques aleatorios, máquinas de soporte vectorial y CN2, fueron

puestos a prueba para obtener la clasificación de ocho especies capturadas con LIDAR batimétrico, luego de la valoración con los diferentes algoritmos llegan a la conclusión que el algoritmo que más ese destaque en la clasificación y de acuerdo con las estadísticas obtenidas fue el algoritmo implementado en la clasificación a partir de bosques aleatorios (Collin et al. 2011).

García et al. (2009) de la Universidad de Sevilla en España, proponen de igual forma una clasificación mediante el empleo de algoritmos como Naive Bayes y C4.5, obteniendo en su clasificación mejor beneficio en la precisión con el uso del algoritmo C4.5 (Garcia et al. 2009).

Todos los autores llegan a conclusiones similares en cuanto a la exactitud en la aplicación de las técnicas de minería de datos en la clasificación de características del terreno de nubes LIDAR, pero no se expone por parte de los autores características acerca del rendimiento computacional de estos algoritmos, donde esta característica si resulta ser propósito del presente artículo.

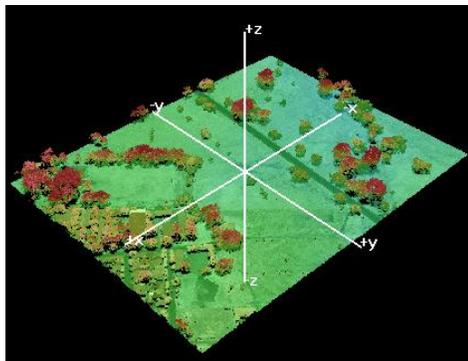
2. Datos y Métodos

2.1. Área de Estudio y Descripción de los Datos

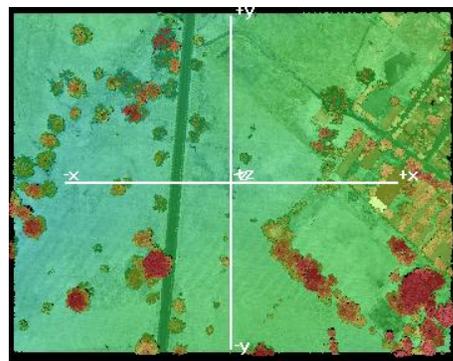
En cuanto a los datos usados para el presente estudio, se utilizó una nube de puntos del sistema de Tecnología LIDAR cubriendo una extensión geográfica de tierra como se muestra en las imágenes *a.* y *b.* de la *Figura 2*, comprendida desde los 676200 m Este y los 857110 m Norte hasta los 676620 m Este y los 857430 m Norte en el sistema de referencia espacial Universal Transversal Mercator (UTM), correspondiente a la Zona 18 Norte.

La extensión cubierta por la nube de puntos hace parte del Municipio de San Alberto en el Departamento del Cesar, ubicado al norte de Colombia. Su clima es cálido, su vegetación muy diversa, tierras fértiles y fauna variada.

Este sistema LIDAR tiene una tasa de pulsos láser de hasta 150 kHz (150 000 pulsos por segundo), un tamaño de huella de 150 a 500 mm, una tasa de escaneo de 10 a 100 líneas por segundo y un ángulo de escaneo de 30°. El archivo *.LAS analizado está constituido por 211131 pulsos. El área de trabajo corresponde con 13, 44 hectáreas. La toma de los datos se realizó el 10 de julio de 2009.



a. Vista Oblicua Área de Estudio



b. Vista Horizontal Área de Estudio

Figura 2. Descripción Zona de Estudio Datos LIDAR

De acuerdo con la clasificación a realizar sobre la nube de datos LIDAR a partir del uso de las técnicas de minería de datos propuesta, se realizó la siguiente clasificación partiendo de las referencias bibliográficas consultadas en donde clasificación más completa encontrada corresponde a la establecida por la Sociedad Americana de Fotogrametría y Percepción Remota/Teledetección (ASPRS), que publicó la última versión en Octubre de 2010.

La clasificación a usar con las técnicas de minería y que compone el área de trabajo correspondió a 7 clases informacionales reconocidas sobre el área de estudio (Ver *Tabla 1*), estas clases fueron generalizadas de una gran variedad de clasificación de cobertura del suelo que se encuentra dentro de la nube de datos en consecuencia con la clasificación expuesta por la Sociedad Americana de Fotogrametría y Percepción Remota/Teledetección (Sociedad Americana de Fotogrametría y Percepción Remota/Teledetección, 2005).

Tabla 1. Clasificación Nube de Datos LIDAR

Clase	Descripción	Numero de Pulsos	%
1	Sin clasificar	200642	95,00
2	Suelo	5981	2,83
3	Vegetación baja	754	0,36
4	Vegetación media	665	0,31
5	Vegetación alta	1605	0,76
6	Edificios	1069	0,51
7	Vías	415	0,23

La información obtenida sobre estas clases fue tomada sobre la nube de datos LIDAR a partir de muestras de entrenamiento como se muestra en la *Figura 2*, este proceso se describe más adelante en la sección *Preparación de los Datos* en la Metodología.

Las variables de estudio que permiten determinar esta clasificación y que componen el grupo de datos dentro de la nube de datos LIDAR se enmarca dentro de la especificación propuesta por la ASPRS (Sociedad Americana de Fotogrametría y Percepción Remota/Teledetección, 2005), y para datos LIDAR dichos datos se encuentran dentro de ficheros conocidos como LAS.

El formato *.las contiene datos binarios que consisten en archivo cabecera, registros de longitud variable y datos de puntos. Se trata de un formato estándar aceptado por las

grandes empresas para almacenar todos los atributos LIDAR. La mayor parte de los fabricantes de sensores LIDAR han aceptado este formato. Almacena los atributos en formato binario, incluyendo tiempo GPS, datos de orientación inercial, x, y, z. La ventaja de este formato es que todos los atributos finales clasificados se almacenan en una tabla de bases de datos, lo que permite a los usuarios manejar un único archivo que permite generar gran variedad de productos y realizar multitud de análisis (García, 2010). Los datos seleccionados para el estudio se encuentran en la *Tabla 2* y se describe a continuación.

Tabla 2. Formato 1 Datos LIDAR para el Estudio

Ítem	Formato	Tamaño	Requerido
X	long	4 bytes	*
Y	long	4 bytes	*
Z	long	4 bytes	*
Intensidad	unsigned short	2 bytes	
Número de Retorno	3 bits (bits 0, 1, 2)	3 bits	*
Números de Retornos (Pulso Emitido)	3 bits (bits 3, 4, 5)	3 bits	*
Bandera de Dirección del Escáner	1 bit (bit 6)	1 bits	*
Borde de Línea de Vuelo	1 bit (bit 7)	1 bits	*
Clasificación	unsigned char	1 bytes	*
Grado del Angulo de Exploración (-90 a +90) - Lado Izquierdo	char	1 bytes	*
Datos de Usuario	unsigned char	1 bytes	
ID Punto Fuente	unsigned short	2 bytes	*
Tiempo GPS	Double	8 bytes	*

Los valores X, Y, Z se almacenan como enteros largos. Se combinan con los valores de escala y de compensación (INS) para determinar las coordenadas de cada punto.

El valor de intensidad (i) es un entero que representa la magnitud del retorno del pulso láser.

Los siguientes cuatro valores se almacenan como campos de byte (García, 2010):

- El retorno (r) es el número de veces que regresan los pulsos. Un único pulso puede tener varios retornos, especificando el orden de retorno según vuelve la señal.
- El número de retornos (n) es el número total de recursos de un pulso. Por ejemplo, los datos de un punto láser puede ser el segundo retorno con un número total de retornos de cinco.
- La dirección de escaneado (d) indica la dirección en la que se mueve el espejo del escáner en el momento de disparar el pulso láser. Un valor de 1 es positivo (de izquierda a derecha) y un valor de 0 es negativo (de derecha a izquierda).
- La línea de vuelo (e) tiene valor 1 cuando el punto está al final del escaneado. Es el último punto antes de que el avión cambie de dirección.

La clasificación (c) representa los atributos de clase de un punto. Si un punto no se clasifica debe ser 0.

El ángulo de escaneado (a) es un campo de un solo byte con un rango entre -90 y $+90$. Es el ángulo en el que el punto láser es disparado desde el avión. El valor 0 representa el nadir, mientras que el valor negativo indica el lado izquierdo del avión. El identificador del punto (p) indica el orden de captura de los puntos. El tiempo GPS (t) es de tipo doble flotante con la información de cada punto. Para este estudio se construyó el modelo de clasificación obtenido c en función de z, i, r, n, a .

3. Metodología

3.1. Preparación de los Datos

Para la preparación de los datos la clasificación de las variables se basó en una selección diferencial de los impulsos de acuerdo con una selección de filtros sucesivos basados en las características de los atributos intensidad, número del retorno, ángulo de escaneo y altura (Fagua et al. 2011). A continuación se explican los filtros empleados en la remoción de atípicos.

El atributo llamado intensidad va de 0 a 255 (debido a que el sistema genera intensidades escaladas a 8-bit), se refiere a la cantidad de energía con que regresa el impulso al sensor de acuerdo a lo expuesto por García et al. (2009) en su trabajo; dicha energía depende del tipo de cobertura con la cual el impulso choca (García et al. 2009).

La intensidad tiene los valores más altos cuando los impulsos rebotan en el suelo y superficies urbanas duras, mientras es menor cuando los impulsos rebotan en vegetación debido a que la clorofila absorbe una fracción grande del infrarrojo (rayo láser) enviado por el sensor, disminuyendo significativamente la energía con que regresa el impulso (Fagua et al. 2011).

Otra variable seleccionada para el análisis fue la altura de los impulsos, la cual debe tener por definición, los puntos más bajos provenientes del terreno (impulsos láser nunca bajan más allá del suelo); en consecuencia, en un lugar relativamente plano, el terreno sería cercano a la línea promedio de los puntos más bajos (Fagua et al. 2011).

Por tal razón a continuación en la *Tabla 3* se analizan las gráficas para las variables altura e intensidad para la remoción de atípicos de la nube de puntos.

Tabla 3. Análisis Estadístico Nube de Datos LIDAR

	z	i	r	n	d	e	a
Mínimo	106.52	0	1	1	0	0	-8
Máximo	140.86	255	4	4	1	0	8
Promedio	113.61	104.8	1.1	1.19	0.5	0	-0.14
Mediana	112.59	123	1	1	0	0	0
Desviación estándar	4.91	51.16	0.32	0.45	0.5	0	3.78
Desviación mediana	1.99	27	0	0	0	0	3
Asimetría	1.8	-0.69	3.46	2.34	0	NA	-0.05
Kurtosis	6.74	2.33	15.34	8.08	1	NA	2.01
Coef. Var. Promedio (%)	4.32	48.81	29.35	38.03	100.07	NA	-2610.85
Coef. Var. Mediana (%)	1.77	21.95	0	0	NA	NA	Inf

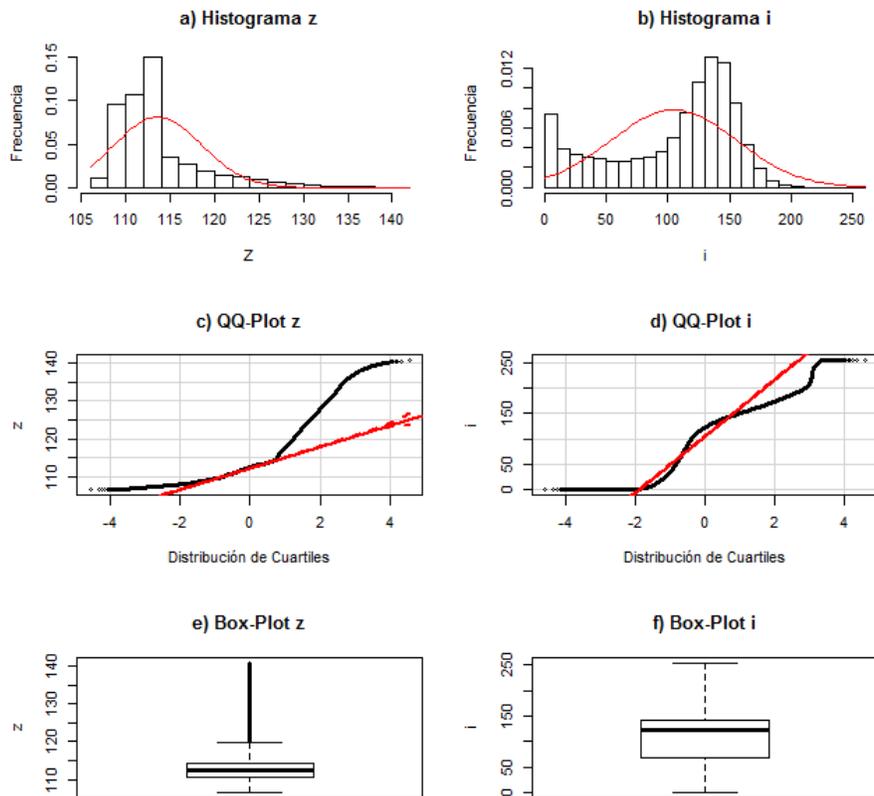


Figura 3. Estadísticas Descriptivas Nube de Datos LIDAR

De acuerdo con las variables analizadas se puede observar en la *Figura 3* se encuentra anomalía en la variable de la altura, lo cual puede ser producto de la toma de ruido correspondiente a aves al momento de captura de los puntos con el láser. Por tal motivo resulta necesaria la remoción de atípicos para la variable z, y para lo cual se empleó el criterio empírico con una heurística de 2 desviaciones en esta remoción, tal y como se muestra en las ecuaciones 1 y 2.

$$\text{Límite superior} = \text{media} + 2 * \text{desviación estándar} = 128.3333 \quad (1)$$

$$\text{Límite inferior} = \text{media} - 2 * \text{desviación estándar} = 98.88209 \quad (2)$$

Por lo tanto con este procedimiento se removieron 4378 puntos que estaban asociados al ruido como se mencionaba anteriormente muy posiblemente asociado a aves que cruzaban por el sector. De acuerdo con esto y en resumen la cantidad de datos a clasificar corresponde con el valor de 196.264 puntos de un total inicial de 200.642 menos los 4378 removidos.

De la muestra total se seleccionaron 6334 (3% pulsos totales) pulso de entrenamiento de manera aleatoria y los 4155 restantes (2% pulsos totales) se emplearon para validación.

Para cubrir las necesidades del aprendizaje supervisado se debe realizar una clasificación de los datos previamente. Esta tarea es muy importante para obtener buenos resultados. De acuerdo a la literatura consultada y al conocimiento de expertos se ha aplicado para clasificar manualmente sobre un 3% de los datos totales (García et al. 2009), se seleccionaron 200 muestras de entrenamiento correspondientes a la categorización de las 7

clases en la nube de puntos. Para efectos prácticos se fue tomada la clase 7 establecidas por la ASPRS correspondiente a Puntos bajos (ruido) como vías.

García (2010) también sugiere clasificar manualmente un 3% de los datos totales para el de entrenamiento del algoritmo. En este trabajo se clasifico manualmente el 5% del total de los datos, siendo una muestra equivalente a 10074 pulsos. Siendo empleado un 3% del total de los datos para entrenamiento y el 2% restante para validación. En la *Tabla 1* se muestra el número de pulsos a clasificar, así como el número de pulsos clasificados manualmente por clase y su proporción sobre el total de pulsos.

En la *Figura 4* se puede observar la selección de la muestra de entrenamiento clasificada sobre la nube de datos LIDAR.

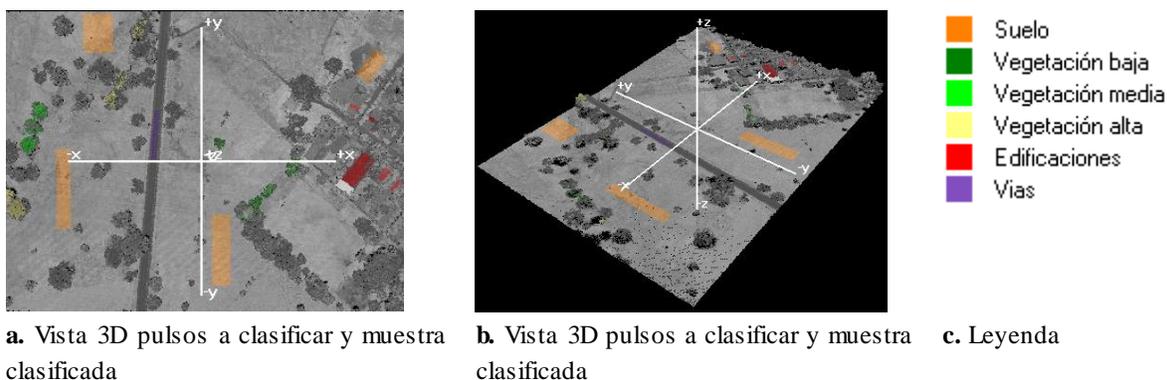


Figura 4. Selección de Muestras de Entrenamiento Clases Definidas

Luego de obtenida la muestra de entrenamiento, ya se tiene una caracterización de las diferentes clases seleccionadas de los puntos LIDAR sobre el terreno, y con lo cual se puede realizar la clasificación de toda la nube. Para esta clasificación se aplicó el árbol bayesiano propuesto por Breiman & Friedman (1984), el cual es el algoritmo contemplado

en la librería *rpart* de *R*, la cual fue usada en la etapa de clasificación de la nube de puntos (Theureau et al. 2014).

3.2. Aplicación del Modelo De Minería Árboles de Decisión

Un árbol de decisión es una forma gráfica y analítica de representar todos los eventos (sucesos) que pueden surgir a partir de una decisión asumida en cierto momento. Nos ayudan a tomar la decisión más acertada, desde el punto de vista probabilístico, ante un definido número de posibles decisiones. Permite desplegar visualmente un problema y organizar el trabajo de cálculos que deban realizarse (Silvente et al. 2013). Los árboles de decisión son una técnica estadística para la segmentación, la estratificación, la predicción, la reducción de datos y el filtrado de variables, la identificación de interacciones, la fusión de categorías y la desratización de variables continuas (Starek et al. 2012).

En cuanto a la evaluación realizada para los datos a partir del algoritmo empleado de árboles de decisión, se comienza por resolver los nodos más cercanos a los resultados finales, en caso de los eventos aleatorios al conocerse la probabilidad se aplica el criterio del valor esperado, en el caso de los nodos de decisión se elige el mejor resultado o el mejor valor esperado, finalmente se llega a la elección de la mejor alternativa. El Software *R* usa un algoritmo de clasificación por árboles de decisión con base a un árbol bayesiano (Theureau et al. 2014).

La inducción resultado de la clasificación del suelo mediante el uso de árboles de decisión de las 7 clases que se usaron de entrenamiento, se obtuvo el siguiente diagrama de árbol (*Figura 5*) de donde se puede ver la distribución de probabilidades realizada por el

algoritmo de clasificación mediante el uso de la librería *rpart* para evaluación de los árboles de decisión sin poda de la librería *rpart* del Software *R* (Therneau et al., 2014).

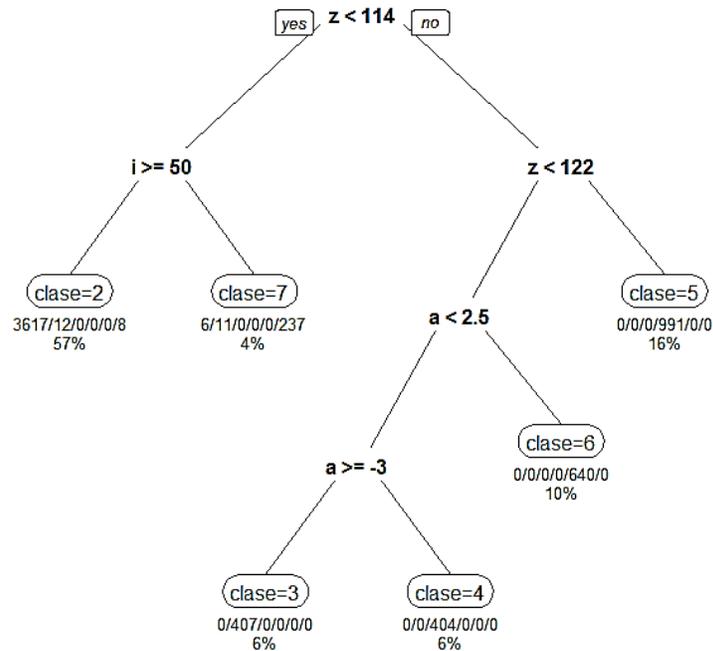


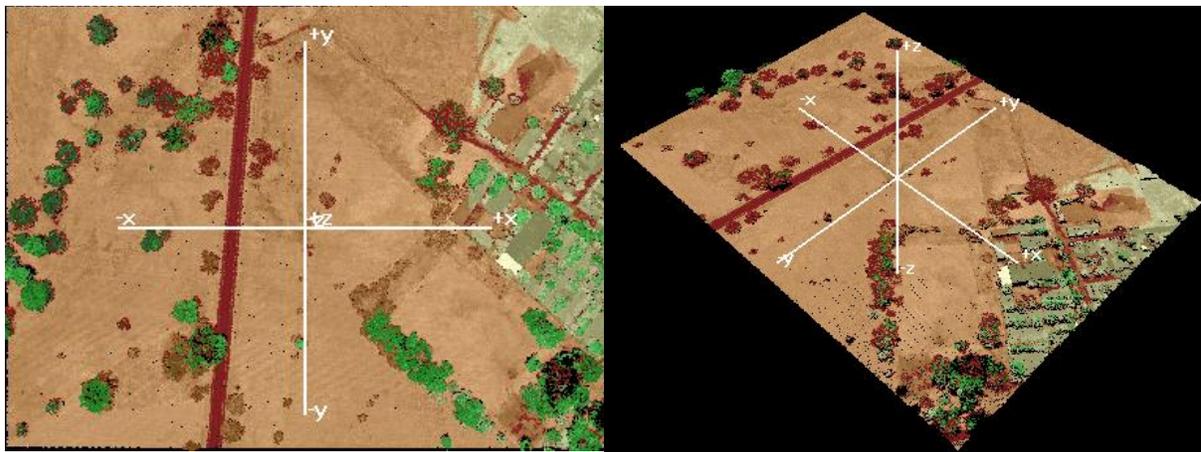
Figura 5. Árbol de decisión clasificación de cobertura sobre la imagen

Del árbol resultado al implementar el algoritmo como se observa en la *Figura 5* se obtiene las siguientes reglas de decisión:

- Si z es menor a 114, entonces se encuentra entre las clases 2 y 7.
- Si el valor de i es mayor o igual que 50 es igual a la clase 2.
- Si i es menor o igual a 50 es igual a la clase 7.
- Si z es mayor a 114, entonces se encuentra entre las clases 5, 6, 4 y 3.
- Si z es mayor a 122 la clase es igual a 5.
- Si z es menor a 122 y a es mayor a 2.5 entonces la clase es igual a 6.

- Si z es menor a 122 y a es menor a 2.5 entonces a es menor o igual a -3 y la clase es igual a 3.
- Si z es menor a 122 y a es menor a 2.5 entonces a es mayor o igual a -3 y la clase es igual a 4.

Luego de que el algoritmo implementa el algoritmo para la clasificación de las clases seleccionadas, se puede observar en la *Figura 6* los resultados obtenidos sobre la nube de puntos LIDAR.



a. Vista 2D Clasificación

b. Vista 3D Clasificación

Figura 6. Clasificación Obtenida de la Nube de Puntos con el algoritmo de Árboles de Decisión

Tal como se muestra en la *Figura 6* y como se puede corroborar en las estadísticas obtenidas (Ver *Tabla 4*) se obtuvo una buena y correcta clasificación de las clases con su correspondencia en el terreno.

Con los datos obtenidos en la validación se construye la matriz de confusión para la clasificación, en la *Tabla 4* se muestra la correspondencia que existe entre ambas fuentes en la clasificación, tanto la tomada en campo como la validación hecha o predicha por el algoritmo, la diagonal expresa el número de puntos de verificación en donde concuerdan

ambas fuentes, mientras que los marginales suponen errores de asignación, por lo tanto se puede observar que la correspondencia de la clasificación obtenida está cercana a la realidad.

Tabla 4. Matriz de confusión asociada a la clasificación mediante Árboles de Decisión

	<i>Sin Clasificar</i>	<i>Suelo</i>	<i>Vegetación Baja</i>	<i>Vegetación Media</i>	<i>Vegetación Alta</i>	<i>Edificios</i>	<i>Vías</i>	<i>Total</i>	<i>ACC Usuario</i>
<i>Sin Clasificar</i>	2302	12	0	0	0	13	26	2353	97,83
<i>Suelo</i>	12	400	0	0	0	0	8	420	95,24
<i>Vegetación Baja</i>	0	0	220	10	12	0	0	242	90,91
<i>Vegetación Media</i>	0	0	15	281	8	0	0	304	92,43
<i>Vegetación Alta</i>	0	0	20	11	403	0	0	434	92,86
<i>Edificios</i>	5	21	0	0	0	147	3	176	83,52
<i>Vías</i>	17	11	0	0	0	11	191	230	83,04
<i>Total</i>	2336	444	255	302	423	171	228	4159	
ACC Productor	99%	90%	86%	93%	95%	86%	84%		

3.3. Análisis de Resultados

Haciendo un análisis gráfico inicial de los resultados obtenidos en la clasificación, se puede observar en la *Figura 6* que existe correspondencia entre los datos clasificados y la información en el terreno.

En cuanto a la exactitud temática evaluada en la aplicación del algoritmo mediante arboles de decisión fue de 94.83 % lo cual indica que más del 90% de los datos se encuentra bien clasificado.

En la *Tabla 4* se observa el porcentaje para cada clase con relación a la clasificación realizada por el usuario versus la clasificación realizada por el algoritmo, donde de forma general se deduce que dentro de las clases con menor correlación se encuentran las edificaciones y el suelo, dado a que tienen una mayor dispersión dentro de la nube de puntos con relación a las otras clases.

La matriz de confusión advierte que la información obtenida de la clasificación supervisada corresponde a una buena clasificación, dada su exactitud temática se acerca al 100 %.

En cuanto al tiempo de cómputo del algoritmo se obtuvo un tiempo de respuesta en la predicción de 1 Segundo, lo cual supera muchos de los programas computacionales usados para este tipo de clasificaciones.

4. Conclusiones

En este trabajo se ha realizado una estimación de la clasificación de la cobertura del suelo existente del área estudiada, se realizó con un número de clases mínima, con el objetivo de comprobar si puede hacerse una evaluación rápida de la superficie mediante la aplicación de algoritmos y técnicas de minería de datos mediante el uso de árboles de decisión.

Aunque a mayor número de clases se discriminan más detalles, ello conlleva a aumentar la probabilidad de cometer errores de clasificación. En cambio, si se tiene el número de clases suficientes se consigue homogeneidad y buena exactitud temática.

Los reportes estadísticos indican la fiabilidad del modelo, al igual que se complementa en el análisis visual de las clases.

También se observa que la mayoría de las clases se encuentran bien definidas dentro de la clasificación de la nube de datos, en donde se resalta la eficiencia del algoritmo es este tipo de estudios.

La exactitud temática mostrada es ciertamente excelente para el bajo tiempo de cómputo empleado en la clasificación.

Y a nivel general se cumplió con el objetivo principal de este trabajo, que era comparar el desempeño ofrecido en la clasificación a partir de la técnica de minería de datos mediante el uso de el algoritmo de árboles de decisión que promovieran buenas exactitudes en la clasificación y alto rendimiento computacional.

Los resultados obtenidos, permitieron ratificar la hipótesis de que la aplicación de estos métodos permite obtener una buena clasificación en aplicaciones de uso y cobertura de la tierra.

5. Referencias Bibliográficas

- Chehata, N., Guo, L., & Mallet, C. (2009). Airborne LIDARFeature Selection for Urban Classification Using Random Forests. In *IAPRS* (Vol. XXXVIII, pp. 207–212). Paris, France.
- Collin, A., Archambault, P., & Long, B. (2011). Predicting Species Diversity of Benthic Communities Within Turbid Nearshore Using Full-waveform Bathymetric LiDAR and Machine Learners. *PloS One*, 6 (6).
- Fagua, J., Campo, A., & Posada, E. (2011). Desarrollo de dos metodologías para la generación de modelos digitales de terreno (MDT) y superficie (MDS) empleando datos LiDAR y programas de licencia. *Análisis Geográfico*, 49, 83–95.

- García, J. (2010). *Remote Mining: Aplicando Minería de Datos a Teledetección Sobre LIDAR*. Universidad de Sevilla.
- García, J., Goncalves, L., & Riquelme, J. (2009). Decision Trees on LiDAR to Classify Land Uses and Covers. Seville, Spain: Universidad de Sevilla.
- Lanig, S., Schilling, A., Stollberg, B., & Zipf, A. (2008). Towards standards-based processing of digital elevation models for grid computing through Web Processing Service (WPS). In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 5073 LNCS, pp. 191–203).
- Lanig, S., & Zipf, A. (2009). Towards Generalization Processes of LiDAR Data based on GRID and OGC Web Processing Services. *Architecture*.
- Niemeyer, J., Rottensteiner, F., & Soergel, U. (2013). Classification of Urban LiDAR Data Using Conditional Random Field and Random Forests. *Urban Remote Sensing ...*, 856(1), 1–4.
- Ocaranza, O., & Quiroz, M. (2005). *Deserci ó n: Huellas Huellas de* (p. 32). Valparaíso.
- Silvente, V., Rubio, M., & Baños, R. (2013). *Cómo Aplicar Árboles de Decisión en SPSS* (Vol. 6, pp. 65–79). Barcelona.
- Sociedad Americana de Fotogrametría y Percepción Remota/Teledetección. (2005). *LAS Specification* (pp. 1–11).
- Starek, M., Vemula, R., & Slatton, K. (2012). Probabilistic detection of morphologic indicators for beach segmentation with multitemporal lidar measurements. *IEEE*, 50(11), 4759–4770.
- Therneau, T., Atkinson, B., Ripley, B., & Ripley, M. (2014). Recursive Partitioning and Regression Trees. CRAN.
- Zhang, J., Lin, X., & Ning, X. (2013). SVM-Based Classification of Segmented Airborne LiDAR Point Clouds in Urban Areas. *Remote Sensing*, 5(8), 3749–3775.