

# EVALUACIÓN DE LA UTILIDAD DE LA MINERÍA DE DATOS PARA LA PLANEACIÓN DE VÍAS PARA EL TRANSPORTE DE CARGA REGIONAL CASO DE ESTUDIO: BOLIVIA

Emma Vanessa Casas Medina, Centro de Estudios Superiores del Estado de Sonora  
Carlos Alberto González Camargo, Universidad Popular Autónoma del Estado de Puebla

## RESUMEN

*El siguiente es un estudio correlacional y explicativo que tiene como objetivo evaluar la utilidad de la minería de datos para la planeación vial del transporte en Bolivia. El ejercicio de minería de datos se realizó a partir de la información generada por el Instituto Nacional de Estadística de este país. La unidad de análisis se limita al estado a partir de la división política de Bolivia, ya que el ejercicio está planteado para la planeación de transporte a nivel regional. Durante el desarrollo del estudio se identificaron clústeres estadísticos regionales, que dadas sus condiciones, son regiones pertinentes para la inversión en infraestructura de transporte que conlleve al mejoramiento de su competitividad. El método utilizado fue en principio la exploración y descripción de la planeación en Suramérica y la utilización de la minería de datos en temas de transporte. Posteriormente se realizó un análisis correlacional mediante la minería de datos, realizando análisis de componentes, factores e identificación de clústeres. Se contrastó esta información con el mapa de vialidad existente y sus características. Finalmente se llegó a la explicación de por qué algunos proyectos no considerados por la iniciativa IIRSA son importantes. Este estudio confirma la potencialidad de la minería de datos para resolver problemas de clasificación orientados a la planeación del transporte. Esta investigación demuestra la posibilidad de utilizar nuevas metodologías para el análisis supraregional del transporte ya que el resultado de los análisis de datos sirve para entender las relaciones de la estadística en el espacio geográfico y su impacto en la planeación del transporte al permitir la identificación de proyectos relevantes que permitan el mejoramiento de la política de infraestructura vial.*

**PALABRAS CLAVES:** Transporte, minería de datos, logística

## DATA MINING USEFULNESS FOR REGIONAL HIGHWAY AND CARGO PLANNING: THE CASE OF BOLIVIA

### ABSTRACT

*In the study, we evaluate the usefulness of data mining for road transportation planning in Bolivia. We conducted a data mining exercise using information generated by the National Institute of Statistics. The study is limited to the states as established by political division of Bolivia. During the development of the study, we identified regional clusters with relevant areas for investment in transport infrastructure that may lead to improved competitiveness. We performed a correlational analysis using data mining. We analyzed the components, factors and identified clusters. We compared this information with the map of existing roads and their characteristics. Finally, we analyzed the reason why some projects were not considered by the IIRSA initiative. We concluded that data mining is a useful tool to solve classification problems in transportation planning.*

**JEL:** F23, R42 R48

**KEYWORDS:** Cargo, Data Mining, Logistic

## INTRODUCCIÓN

En los últimos años el tema de la integración regional ha sido muy importante a nivel global. En el mundo los países se han interesado por construir proyectos de infraestructura vial interregionales con la intención de elevar su competitividad. En Europa y en Estados Unidos las vías para la integración económica y el transporte de mercancías cumplen un papel fundamental para el desarrollo. La Comunidad Económica Europea se ha planteado la construcción de la red transeuropea de transporte con el fin de fomentar la intermodalidad en el movimiento de carga y realizar proyectos de infraestructura conjuntos entre los diferentes países. En Suramérica la más reciente iniciativa nace con el proyecto de Integración de la Infraestructura Regional para Suramérica (IIRSA) liderado por el Banco Interamericano de Desarrollo BID. Este programa pretende integrar el hemisferio sur de América a través de varios proyectos de infraestructura. Al revisar los planes de desarrollo de los países de Suramérica, no se encuentra una propuesta clara de vías para la integración supraregional, ni tampoco ejercicios de modelación de transporte a este nivel que pretendan al mejoramiento logístico. El modelo clásico de transporte, conocido como el de las cuatro etapas, es poco aplicado a nivel nacional en los países de la región.

Lo anterior trae como consecuencia que las decisiones sobre la construcción de nueva infraestructura de transporte no estén basadas en criterios sociales o económicos y con el rigor técnico deseable por parte de los modeladores de transporte, sino que en muchos casos son decisiones de carácter político que responden a sectores específicos de alguna región, representando este ejercicio una aproximación al proceso de planeación. Por esta razón se vuelve inadecuada la red de transporte desde el punto de vista logístico. Esto trae como resultado que de alguna manera los tomadores de decisiones crean que un proyecto puede estar relacionado con el desarrollo básicamente por su apreciación personal. Se encuentran titulares en los diarios que dicen cosas como que la ciudad o en otros casos la nación, tomó la decisión de construir un proyecto, cuando las decisiones realmente las toman las personas y no por un estudio y mucho menos la utilización de la tecnología.

En este contexto, es pertinente analizar la posibilidad de aplicar las técnicas de la minería de datos al análisis de los proyectos de transporte a un nivel internacional impactando la logística y la competitividad. (Ian y Eibe, 2005) Este problema podría resolverse construyendo una base de datos interregional que contenga estadísticas económicas, políticas, sociales y culturales. Posteriormente los resultados obtenidos mediante la minería de datos pueden cruzarse con un sistema de información geográfico que permita la interacción entre estos dos componentes.

Este documento tiene como objetivo demostrar la aplicabilidad de la minería de datos a la planeación del transporte a un nivel supranacional. Según Coro (2008), esta técnica se ha utilizado hasta ahora para el análisis de temas relacionados con los negocios y en algunos casos se ha complementado con el análisis espacial. El trabajo que se plantea requiere utilizar los datos estadísticos generados por el Instituto Nacional de Estadística de Bolivia, ya que este es el país que se ha tomado como caso piloto, con el fin de realizar un ejercicio de minería de datos, que contribuya de alguna manera a la identificación de zonas geográficas que necesitan proyectos de transporte para incrementar su dinámica económica y de esta manera mejorar su competitividad. Además, pretende utilizar las estadísticas socio-económicas como insumo principal para la toma de decisiones sobre infraestructura regional para la integración.

El objetivo de este artículo es demostrar mediante un ejercicio piloto de minería de datos que se puede generar nueva información para Bolivia utilizando la asociación por factores, componentes y clústeres. Esta información permitiría interactuar con el software Geoda y posteriormente comparar los datos con los del sistema cóndor de la Corporación Andina de Fomento. A partir de éstos resultados se podrían realizar recomendaciones de localización de proyectos de transporte para la integración y el mejoramiento logístico para la competitividad.

Esta investigación está desarrollada de la siguiente manera: en la primera parte se expone el Estado del Arte, donde se basa la parte teórica del proyecto a través de la revisión literaria relevante para análisis del caso, describe la metodología que se utilizó, así como las herramientas utilizadas para el desarrollo del mismo, posteriormente se plantean los resultados y por últimos las conclusiones que se llegaron con el desarrollo de la investigación.

## **REVISION LITERARIA**

Hasta ahora la manera convencional de realizar los estudios de transporte es mediante aplicación de encuestas y modelamiento con software especializado específico para el sector transporte como el EMME/2, AISUM2 y GETRAM (Montero et al., 2010). En el caso de las ciudades es posible llegar a utilizar este tipo de software para modelar flujos de pasajeros. Sin embargo, no se hace lo mismo a nivel de planeación regional en Suramérica, para modelar carga que se mueve entre países. Esto ocurre porque las encuestas para suministrar los datos serían muy costosas. Existe la información estadística de cada país generada por el instituto de estadística respectivo. Esta puede ser valiosa si se encuentra la manera de aplicarla a la planeación del transporte regional. Es importante considerar que en el mundo recientemente se han empezado a utilizar las técnicas de la minería de datos para aplicarlas al transporte.

El trabajo de Kelly (2006), muestra las posibilidades de innovación que existen a nivel del Transporte en la era de la información. Aquí se presenta el proyecto MIDAS (Motorway Incident Detection and Automatic Signalling), que tiene como objetivo el control de la señalización de las carreteras en tiempo real según los incidentes, incluyendo la política de Transporte multimodal directo, así como las Isocronas de tiempos de espera generadas con sistemas de información geográfica.

El Wigglybus que no tiene ruta, siendo esta determinada por los usuarios que llaman a un call center. Estas aplicaciones han permitido la construcción de bodegas de información que permiten el desarrollo, implementación y monitoreo de políticas de transporte. Carvalho (2007), ha utilizado la minería de datos para determinar demanda y programación de operaciones en sistemas de trenes. Este es un trabajo que se complementa con el de Bergougnoux (2000), Shawe, Tijl y Cristianini (2006) quienes han escrito sobre la utilidad de los sistemas geográficos de información para el desarrollo de sistemas inteligentes de transporte. Il-Gyo, Susan, y Chi-hyuck (2007) han demostrado en su documento que la minería de datos puede llegar a optimizar procesos brindando mayor calidad.

Por otra parte Kumar (2003), Haluzová (2008), Deer-Horng, Shin-Ting, Chandrasekar (2004) y Kamal (2004), han utilizado las técnicas de la minería de datos para aplicarlas en ingeniería de transporte. La aplicabilidad ha sido principalmente en el estudio de accidentalidad de vehículos y la gerencia del tráfico. Este ejercicio ha permitido el análisis de los flujos de tráfico disminuyendo los costos de la planeación del transporte.

El trabajo de Guisande y Subirada (2001) evidencia que los datos censales pueden ser manejados por medio de las técnicas de la minería de datos para encontrar asociaciones espaciales. En este sentido la utilidad de los datos estadísticos de los departamentos de los países de Suramérica puede cumplir el mismo objetivo. Stolzer y Halford (2007) han realizado ejercicios de minería de datos aplicados a las operaciones aéreas, comparando los resultados frente a los análisis estadísticos convencionales, así como Hian (2004) analiza y compara la utilidad de las redes neuronales, árboles y regresión logística, para predecir una situación de continuidad en el área de transportes. Según lo encontrado por Sheng, Dianhai y Hongsheng (2010) en un experimento realizado en la ciudad de Beijing, la gerencia del tráfico puede ser administrada analizando las relaciones entre volumen, ocupación y capacidad, a partir de una base de datos donde se aplican modelos de estimación utilizando la estadística.

Estos datos estadísticos pueden relacionarse con lo planteado por Krugman (1992, p. 17), donde sugiere que la concentración geográfica nace básicamente de la interacción de los rendimientos crecientes, los costos de transporte y la demanda. Cohen (2010) demostró mediante un análisis de econometría espacial la relación entre la productividad y la inversión en infraestructura de transporte. Si las economías de escala son suficientemente grandes, cada región prefiere abastecer el mercado nacional desde una única región. Para minimizar los costos de transporte, elige una ubicación que permita contar con una demanda local grande. Por último, el modelo de Krugman permite suponer que la demanda de cada bien es estrictamente proporcional a la población del lugar donde es producida. El trabajo de Clifton y Scuderi (2001) permite comprobar la aplicación de la estadística para el análisis del uso de la tierra y el transporte.

Estas experiencias evidencian que en el mundo se han venido desarrollando una serie de proyectos relacionados con el modelamiento de bases de datos, sistemas de información geográfica y datos de transporte, donde finalmente confluyen la teoría económica, el transporte y la minería de datos espacial. De acuerdo a los ejemplos anteriores, no se encuentra evidencia de la utilización de los datos estadísticos generados por los departamentos de estadística, para ejercicios de planeación de transporte a nivel supraregional. Aplicando las técnicas de la minería de datos es posible realizar análisis espaciales con el fin de encontrar la localización de los proyectos de transporte necesarios para el mejoramiento de la dinámica económica regional.

El análisis de transporte a realizar se ocuparía entonces, en el marco de esta dinámica, de la identificación de clústeres regionales y la satisfacción de sus necesidades de infraestructura apropiadas. En un marco de competencia global, este análisis resulta un aspecto clave no sólo para la expansión y el crecimiento del comercio internacional entre países, sino incluso para garantizar su propia supervivencia. Debido a que el territorio de los países y los estados de cada uno de ellos tienen una localización en el espacio, la consideración de esa componente espacial asociada a datos estadísticos resulta fundamental para la planeación de proyectos de transporte. Este tipo de análisis debe responder a preguntas como ¿Dónde se localizan los estados más desarrollados y cuáles son sus características? ¿Dónde se encuentran los sectores de producción? ¿Dónde existe una concentración importante de infraestructura? ¿Dónde debemos proponer la construcción de proyectos para la conexión regional? ¿Dónde se localizan los puntos críticos de la infraestructura? ¿Dónde podemos establecer nuevos puntos de oferta de vías?

## **METODOLOGÍA**

El ejercicio que se pretende realizar es un estudio de tipo descriptivo, correlacional y explicativo, que permitirá inferir juicios sobre los espacios geográficos que necesitan fortalecer su infraestructura de carreteras. Los resultados se constituirán en una herramienta para la toma de decisiones, todo esto con un valor agregado: la espacialidad de los resultados.

La metodología a emplear es la de los autores Hernández et al. (2010) que consta de los siguientes pasos:

1. Alcances y enfoque de la investigación
2. Hipótesis
3. Diseño de la investigación
4. Selección de la muestra
5. Recolección de datos
6. Resultados: Análisis de datos

En cuanto al alcance, la investigación es de carácter mixto, en principio es cuantitativa por ser correlacional, ya que se plantea que existen estados más desarrollados que otros en Bolivia y para esto se utilizará la minería de datos, para encontrar relaciones mediante el análisis de componentes, el análisis de

factores y diferencias entre grupos mediante los dendogramas y clústeres. La investigación es de tipo transversal porque los datos se recolectarán en un único momento.

Finalmente todo el procedimiento termina en un tipo de investigación explicativa porque se comprobará la hipótesis de que a partir del ejercicio correlacional y la información sobre las vías, que es de tipo cualitativo, es posible identificar proyectos importantes para el desarrollo regional, aprovechando las estadísticas existentes y sin modelar mediante software, evitando incurrir en altos costos por levantamiento de información y licencias de programas. La hipótesis que se plantea es que la minería de datos permite identificar diferencias y similitudes entre los estados de Bolivia y establecer relaciones multivariadas entre datos estadísticos existentes. Las diferencias encontradas entre los estados comparados, son un indicativo de cuáles vías son importantes para el mejoramiento de la competitividad, ya que especifican a favor de qué grupo es la diferencia. El análisis de factores y de componentes permite identificar la importancia de las variables según su variabilidad y su relación con las otras variables de la base de datos en una relación de dependencia que sólo podrá identificarse una vez sea realizado el ejercicio de minería de datos.

Como el ejercicio de minería de datos se basa en la estadística descriptiva multivariada los resultados presentados buscaron tres objetivos. Primero, realizar un análisis de factores con el fin de lograr la reducción de un gran número de variables a unos cuantos factores. Este método indica cuántas dimensiones integran a una variable y qué ítems forman cada dimensión. Estos representan combinaciones lineales de las variables analizadas explicando cada uno, en cierto porcentaje, la variabilidad de los datos. Al cumplirse este objetivo es posible identificar para cada factor principal las variables asociadas más importantes según su coeficiente de correlación en la ecuación lineal.

Estas variables indican la diferencia del clúster de clasificación para cada estado. Otros elementos estadísticos a tener en cuenta son la asimetría de la distribución muestral valorada por la curtosis, la forma como se encuentran distribuidas la media, la mediana y la moda, la variabilidad, la regresión lineal entre variables, los coeficientes de correlación y la prueba t. La varianza es una medida que permitirá descartar variables cuya variabilidad sea menor a dos, debido a que este dato no representaría un factor diferenciador entre departamentos sino más bien sería sinónimo de homogeneidad.

El diseño será concurrente porque al tiempo que se avanza en la investigación cualitativa sobre planeación del transporte relacionada con la forma como se hace en el mundo, en Suramérica, en Bolivia y lo que plantea el IIRSA, al mismo tiempo se avanza en la recolección de la información para la conformación de las bases de datos a nivel de estado para Bolivia con el fin de realizar el ejercicio de minería de datos. Al final se integra el ejercicio de minería de datos con la propuesta del IIRSA en la zona de Bolivia mediante la comparación y la explicación de proyectos viales que deberían construirse.

El país seleccionado para esta prueba piloto es Bolivia por la disponibilidad de información. La unidad de análisis es el estado. Para cada uno de los estados se recolectó la información necesaria para realizar el ejercicio de minería de datos. La fuente de información fue el Instituto Nacional de Estadística de Bolivia. De aquí se seleccionaron las variables socio-económicas a criterio de los investigadores basado en lo que se ha utilizado en estudios de transporte, por lo cual es no probabilística. Los requerimientos de información para la planeación de un proceso de transporte están constituidos por los datos georreferenciados sobre vías existentes, características de las vías y datos socio-económicos (Ortúzar, 1994). Los datos sobre las vías existentes se tomaron del sistema de información geográfica Cándor de la Corporación Andina de Fomento. Esta capa incluye las características de las vías regionales. Para cada uno de los países se conformará una bodega de datos con variables económicas y sociales.

Las variables utilizadas fueron cifras del censo 2001 de los indicadores de analfabetismo, promedio de años de estudio, asistencia escolar, máximo nivel educativo, población en edad de trabajar, población en

edad de no trabajar, población ocupada, población económicamente activa, población económicamente inactiva, tasa bruta de la población, tasa de oferta potencial, tasa global, tasa de dependencia, pobreza, inadecuados materiales de construcción, inadecuados servicios de agua y saneamiento, insuficiencia en educación, inadecuada atención en salud y energía eléctrica.

Datos del año 2006 relacionados con la deserción inicial, deserción primaria, deserción secundaria, relación alumno-docente, crecimiento de la población universitaria y producto interno bruto. Cifras del año 2007 relacionadas con los establecimientos de salud, producción en el área de la agricultura, ganadería, caza y silvicultura, industria manufacturera, comercio, transporte, almacenamiento y telecomunicaciones, construcción, educación, servicios sociales y de salud, alimentos y bebidas no alcohólicas, vivienda, agua, electricidad, gas, combustibles, transporte, prendas de vestir y calzados, restaurantes y hoteles, producción de cemento, consumo de agua, flujo de pasajeros de transporte aéreo, flujo de carga aérea, exportaciones de gas, zinc, plata, pieles, uso del modo férreo, carretero, aéreo, ductos, importación de equipos de transporte, inversión extranjera directa, inversión productiva, inversión en infraestructura, inversión social, cartera, exportaciones a Brasil, Argentina, Perú, Colombia, Ecuador y Venezuela. Proyecciones para el año 2008 de población, tasa de crecimiento de la población, población por género, superficie, densidad, esperanza de vida, tasa de mortalidad infantil y edad promedio.

Para la recolección de datos se construyó un base de datos de las variables relacionadas para cada estado de Bolivia a partir de la información del Instituto Nacional de Estadística de este país, buscando los datos más actualizados. Este archivo están conformado por un número de filas igual al número de departamentos que es de nueve y un número de columnas suficiente para albergar todas las variables. Una vez construido este archivo se realizaron los cálculos de minería de datos y los procedimientos para el análisis de resultados según como lo sugiere Flórez (2005). Las capas de información fueron obtenidas del sistema geográfico cóndor de la Corporación Andina de Fomento CAF y se conectaron a la tabla de Excel diseñada para cada país. Para el ejercicio de minería de datos se utilizó la extensión XLSTAT, del software Excel que permitió llevar a cabo el análisis de estadística descriptiva, el análisis de factores, el análisis de componentes, la construcción de dendogramas y la identificación de clústeres siguiendo un procedimiento similar al de Stolzer y Halford (2007).

## RESULTADOS

Para la caracterización de los departamentos de Bolivia se realizó un primer análisis de los datos para reducir la matriz de información. Aquí se descartaron las variables que tenían una varianza menor a 2 porque los datos al ser semejantes no se constituyen en factores diferenciadores del desarrollo de los departamentos. Estas variables fueron el promedio de años de estudio, la tasa de dependencia y las exportaciones a Ecuador. Igualmente se eliminaron las variables que para el estudio de componentes principales tenían signo negativo en tres de ellos. Esto quiere decir, que en la ecuación lineal del componente no están relacionados positivamente, sino negativamente, o sea no tienen relación positiva con el factor, por lo tanto no aportan a él, no tienen relación.

El análisis de componentes principales tiene como objetivo la síntesis de la información o la reducción de la dimensión (número de variables). Es decir, de la base de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible. Los nuevos componentes principales o factores serán una combinación lineal de las variables originales, y además serán independientes entre sí. Con la nueva matriz, se redujo el número de variables de 76 paso a 54.

Utilizando el análisis de componentes principales se redujo este juego de datos a cuatro factores. Como se presenta en la Tabla 1, los valores que son identificados con los nombres de F1, F2, F3 y F4, están constituidos por combinaciones lineales de las variables estudiadas. La agrupación la hace el software de minería de datos basado en las variabilidad de la información. Es decir, cuando se hace la agrupación por

dendogramas de los departamentos, es porque la varianza de los datos en la base es parecida. Al ser similar entonces se definen los diferentes grupos o clusters que permiten identificar la similitud entre departamentos.

Tabla 1: Valores Propios Análisis de Factores Bolivia

	F1	F2	F3	F4
Valor propio	22,099	12,451	11,309	8,141
Variabilidad (%)	40,924	23,057	20,942	15,076
% acumulado	40,924	63,982	84,924	100,000

Resume los cuatro factores en los que puede ser representados todo el conjunto de variables. Lo que interesa es la variabilidad. Por ejemplo para el factor F1 en la tabla se muestra que representa el 40,9%. O sea que solamente considerando la combinación lineal de este factor tendrías una representación del conjunto de datos de este porcentaje.

Fuente: Ejercicio de minería de datos

Como se detalla en la Figura 1, el primer componente es el que tiene la varianza más alta y por lo tanto explica en un mayor porcentaje los datos. Este factor alcanza una variabilidad del 40,92%. En la Tabla 2 se han ordenado de mayor a menor las variables según el coeficiente de cada una de ellas en la ecuación del factor. En este primer componente se evidencia la importancia de las exportaciones hacia Perú, ya que esta variable tiene un coeficiente de correlación de 0,204 en la ecuación lineal del factor.

Tabla 2: Vectores Propios

Variable	F1
Perú	0.204
X Zinc	0.204
Venezuela	0.203
I productiva	0.203
Cartera	0.203
M equipos de transporte	0.199
Consumo agua	0.199
Carretero	0.199
Población	0.198
X pieles	0.197
Servicios sociales y de salud	0.196
PIB	0.193
PET	0.189
Educación	0.188
PENT	0.181
Modo férreo	0.180

Se muestran los coeficientes de correlación de cada variable con el F1. 0,204 X Perú identifica las exportaciones hacia este país. Puede armarse una ecuación de cada componente (cada F) con los resultados, de tal manera que para F1 se tiene que la ecuación es:

Componente 1 = 0,204 X Perú + 0,204 X zinc + 0,203 X Venezuela + 0,203 I y así se suman todas las variables mostradas en la tabla 2

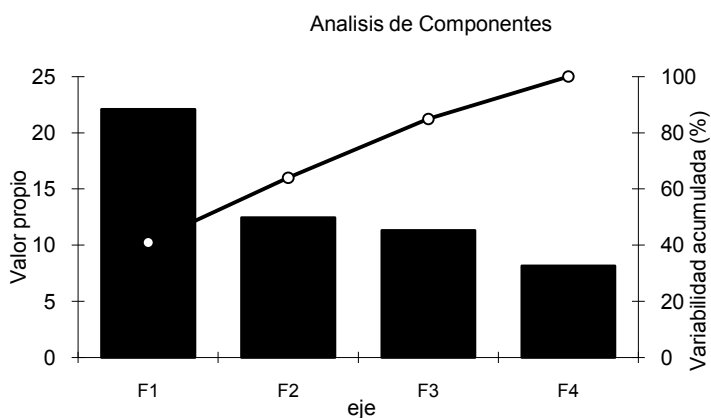
Fuente: Ejercicio de minería de datos

Otras variables importantes son las exportaciones hacia Venezuela, las exportaciones de zinc, las de pieles, la inversión productiva, la cartera, la importación de equipos de transporte, el consumo de agua, el crecimiento del modo carretero, la población, los servicios sociales y de salud, el PIB, la población en edad de trabajar, el crecimiento de la educación y el crecimiento del modo férreo, entre otras variables. Este componente está asociado entonces principalmente a las exportaciones, la población, el PIB y los modos carretero y férreo. La ecuación de este componente según los vectores propios para el factor 1 es: Componente 1 = 0,204 X Perú + 0,204 X zinc + 0,203 X Venezuela + 0,203 I

El segundo componente explica un 23,06% la variabilidad total. Las características de este factor están relacionadas con la carga aérea, el analfabetismo, las exportaciones a Argentina, Colombia y Brasil. Está relacionado también con la insuficiencia en educación, las exportaciones de gas, los restaurantes y hoteles, la densidad, la producción de cemento y las prendas de vestir y calzados.

El tercer componente explica el 20,94% de la variabilidad total, completando hasta aquí una variación acumulada del 84,92%. Las características de este factor están relacionadas con el crecimiento de la población universitaria, la tasa de oferta potencial, las exportaciones de plata, la deserción en primaria, el crecimiento del grupo de transporte, almacenamiento y comunicaciones, la edad promedio, el crecimiento de la producción de prendas de vestir y calzados, la pobreza, el transporte y el modo férreo. Nuevamente encontramos la participación del transporte como uno de los factores diferenciadores.

Figura 1: Variabilidad Acumulada por Factores, Bolivia



Expone un comparativo de las varianzas siendo el componente F1 el que mayor, alcanzando una variabilidad del 40.9%, el segundo componente presenta una varianza del 23.06%, el tercero presenta una variabilidad del 20.94%, completando una variación acumulada del 84.92%  
Fuente: Análisis de factores en Excel

Ahora se realiza el análisis de clúster para identificar perfiles de departamentos similares entre sí. En el dendograma mostrado en la Figura 2 se identifican asociaciones entre los departamentos.

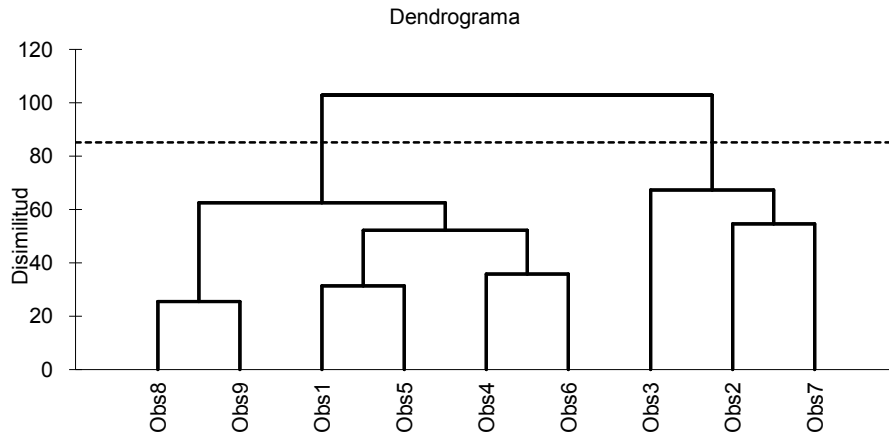
De aquí se concluye que para los nueve departamentos de Bolivia se pueden realizar dos agrupaciones según las características de sus variables. El primer clúster está conformado por Chuquisaca, Oruro, Potosí, Tarija, Beni y Pando, el segundo clúster por la Paz, Cochabamba y Santa Cruz. La Figura 3 muestra los proyectos de IIRSA en Bolivia. El proyecto de conexión con Perú es Puno - La Paz. En el interior tenemos los tramos Potosí - Sucre y Potosí - Tarija.

El proyecto de conexión con Perú en el tramo Puno - La Paz, conecta nuestros departamentos identificados como clúster 1 de Bolivia con Perú. Por lo cual se deduce que el proyecto es importante para la región.

El transporte es uno de los factores que dinamiza el desarrollo. Para esto deben buscarse las conexiones viales del clúster 2 al clúster 1 según población y crecimiento del PIB. Los tramos Potosí - Sucre y Potosí - Tarija buscan este objetivo según población y PET.



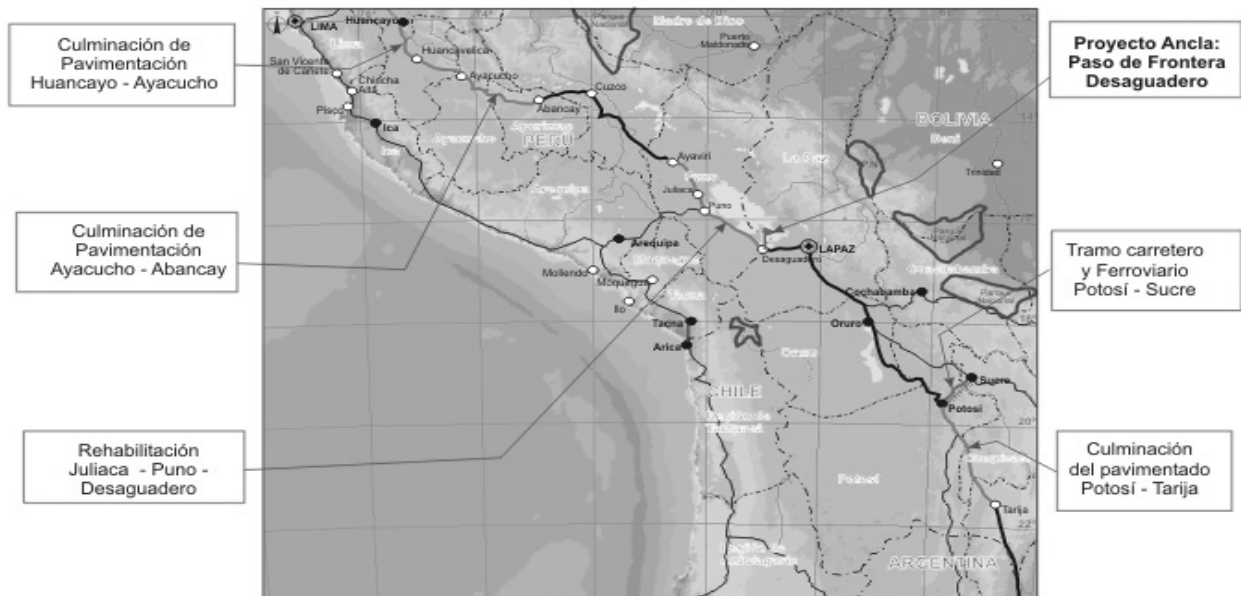
Figura 2: Dendogramas, Bolivia y Sus Respectivos Departamentos de Bolivia



Obs1	Obs4	Obs5	Obs6	Obs8	Obs9	Obs2	Obs3	Obs7
CHUQUISACA	ORURO	POTOSÍ	TARIJA	BENI	PANDO	LA PAZ	COCHABAMBA	SANTA CRUZ

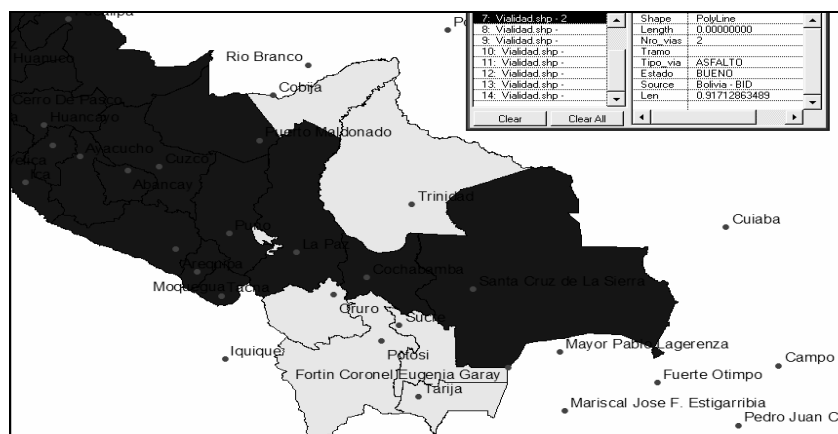
Se muestran los proyectos de IIRSA en Bolivia la cual se divide en dos agrupaciones según las características de sus variables el primer clúster está conformado por Chuquisaca, Oruro, Potosí, Tarija, Beni y Pando, el segundo clúster por la Paz, Cochabamba y Santa Cruz. Fuente: Ejercicio de minería de datos.

Figura 3: Grupo 8 Eje Andino- Conexión Peru-Bolivia (Huancayo-Ayacucho-Tarija-Bermejo)



En esta figura se muestra que el proyecto de conexión con Peru es Puno – la Paz y en su interior sus puntos de conexión son Potosí,- Sucre y Potosí – Tarija. Fuente: IIRSA

Figura 4: Zona Límite Entre Perú y Bolivia



La Figura muestra los límites entre Perú y Bolivia, así como los departamentos de cada una de ellas. Fuente: mapas en Arcview a partir de fuentes de datos sistema cónдор 3.0

Sin embargo el aporte que se puede dar a partir de lo obtenido en este trabajo, es que por crecimiento del PIB debería existir un proyecto para mejoramiento de la vía Oruro – La Paz que en parte de su trayecto está regular según los datos de estado de la vía de la capa de viabilidad en Arcview.

La conexión internacional entre Perú y Bolivia como se muestra en la Figura no 4, estaría dada por los tramos Río Branco-Cobija-Puerto Maldonado-Cuzco. Por lo cual el proyecto Río Branco – Cobija debe ser complementado con Puerto Maldonado – Cuzco, ya que esta vía de Cobija a Puerto Maldonado es un sendero, lo que significa que no está en buen estado de pavimentación. Si no se complementa se pierde la continuidad de la carretera, tendríamos el primer tramo pavimentado, pero el segundo en condiciones difíciles de tránsito. Es por esto que se concluye que aunque el proyecto IIRSA considere la inversión en el tramo de Puerto Maldonado a Cuzco y Juliaca, debe antes arreglarse Cobija – Puerto Maldonado, ya que en la red esa vía aparece como un sendero, de otra manera este tramo quedaría en mal estado y es el intermedio entre Río Branco y Puerto Maldonado. Existe otra posibilidad interesante de inversión para la integración, por la reducción del costo logístico de transporte, concebido como un proyecto multimodal para transportar carga por el río Madre de Dios hasta La Paz. Lo anterior permitiría la conexión Porto Velho – La Paz. Esto se evidencia en la salida espacial o mapa a partir de la superposición de las capas de cluster con la fluvial, por simple inspección.

## CONCLUSIONES

Esta investigación se centró en la aplicación de la minería de datos o estadística descriptiva multivariada a la clasificación de los estados de Bolivia. Con base en esta clasificación se examinaron utilizando sistemas de información geográfica, los proyectos planteados por la Integración de la Infraestructura Regional para Suramérica IIRSA, con el fin de confirmar la necesidad de los proyectos e identificar otras vías importantes para el desarrollo en este país, esta es la principal contribución ya que no se había realizado con anterioridad un ejercicio de minería de datos con este fin, generalmente se había utilizado para toma de decisiones en entidades financieras, en este trabajo se utilizó esta técnica para la planeación del transporte, aprovechando las bases de datos del instituto de estadística de Bolivia, se generó nueva información a partir de la ya existente.

El caso estudiado, confirma la potencialidad de la minería de datos para resolver problemas de clasificación de utilidad para la planeación del transporte. Adicionalmente ha puesto de manifiesto la importancia de la información generada por los departamentos de estadística de los diferentes países, ya que estos datos se constituyen en variables relevantes para el análisis descriptivo multivariado. El análisis

de series geográficas requiere de herramientas propias que van más allá de las convencionales técnicas estadísticas, como son los sistemas de información geográfica. Como ejemplo se utilizaron en este trabajo el Geoda, el Arcview y la información del sistema cóndor de la CAF. Estas herramientas deben estar dirigidas al análisis de la asociación, dependencia o concentración espacial. El análisis sirve como apoyo a la toma de decisiones sobre que ruta o rutas son las que se deben mejorar para incentivar el transporte entre países y aumentar la competitividad regional. En el caso del transporte y en ausencia de ejercicios de modelación suprarregional, la identificación de clústeres contribuye al mejoramiento de la planeación de vías para la competitividad. En todo caso, más allá del interés que pueda derivarse de los hallazgos de esta investigación, la misma debe interpretarse como una contribución al análisis de la planeación del transporte a nivel de países.

A partir de este ejercicio se puede mejorar y entender los impactos de la estadística en el espacio geográfico y su impacto en la planeación del transporte. El resultado de los análisis de datos puede servir para mejorar las políticas en materia de transporte. La principal dificultad para pensar en un aplicativo de orden internacional es la consecución de las bases de datos, el acceso a la información completa y la normalización de la misma. El modelo se puede utilizar para planeación del transporte a nivel país pero estará limitado por la información que generan los departamentos de estadística. Esta información suele ser diferente de un país a otro, por lo cual los clusters estadísticos encontrados no tendrán las mismas características, razón por la cual no son comparables. Este modelo es útil en el caso de planeación a nivel macro pero está limitado para la planeación a nivel ciudad, ya que para este caso es necesario elaborar un estudio de origen destino para evaluar la oferta y demanda de transporte.

En estudios futuros se recomienda el tener contactos a nivel internacional que puedan revisar y validar en campo la potencialidad de las vías seleccionadas como proyectos para la competitividad regional. Esto confirmará que las conclusiones a las que se ha llegado utilizando las diversas herramientas son confiables. Este ejercicio puede realizarse para los diferentes países con el fin de identificar los proyectos importantes e incluso más allá del estudio del transporte podrían realizarse investigaciones que utilicen la técnica de la minería de datos para tomar decisiones a nivel de mercados internacionales, salud, educación o cualquier otra variable importante para la sociedad.

## REFERENCIAS

- Bergougnoux, P. (2000). Geographic Information Systems and Intelligent Transportation Systems. *Geoinformática*. 4(2), 123-125.
- Blythe, P. (2007). Future intelligent infrastructure: Transport in 2050. *Paper Presented at the Sage Faculty Seminar. Octubre 30, 2007*. Reino Unido.
- Carvalho G. (2007). Methodology for Railway Demand Forecasting Using Data Mining. *Paper Presented at the SAS Global Forum 2007. Abril 16, 2007*. Orlando, Estados Unidos.
- Cohen, J. (2010). The broader effects of transportation infrastructure: Spatial econometrics and productivity approaches. *Transportation Research Part E: Logistics and Transportation Review*. 46(3), 317-326.
- Coro, Y. (2007). El Geomarketing y la distribución comercial. *Investigación y marketing*. 79, 6-13.
- Der-Hong, L., Shin-Ting, J. y Chandrasekar, P. (2004). Applying data mining techniques for traffic incident analysis. *Journal of the institution of engineers*. 44(2), 90-102.

Esser, Klauss. (1996). *Competitividad sistémica: un nuevo desafío para las empresas y la política*. Revista de la Cepal. 59, 39-52.

Flórez, J. (2005). Introducción al Data Mining. *Documento presentado en el seminario sobre Geomarketing en 2005 en el Institute for International Research. Noviembre-Diciembre, 2005*. Madrid, España.

González, C. (2009). *Sistemas de Información Geográficos para los negocios internacionales*. Bogotá : Universidad La Gran Colombia.

Guisande, C. y Subirada, Francesc. (2001). El cens de població: un assaig d'interpretaci'o mitjançant data mining. *Questiío*. 25(3), 553-580.

Haluzová, P. (2008). Effective Data Mining for a Transportation Information System. *Acta Polytechnica*. 48(1), 24-29.

Hérrnandez Sampieri, R., Fernández Collado, C. y Baptista, P. (2006). *Metodología de la investigación* (4ª ed.). México: Mc Graw-Hill.

Hian Chye Koh, Chan Kee Low, (2004) "Going concern prediction using data mining techniques", *Managerial Auditing Journal*, Vol. 19 Iss: 3, pp.462 - 476

Huang, L., Stinchcomb, D., Pickle, L., Dill, J. y Berrigan, D. (2009). Identifying Clusters of Active Transportation Using Spatial Scan Statistics. *American Journal of Preventive Medicine*. 37(2), 157-166

Ian W y Eibe F (2005). *Data Mining. Practical Machine Learning Tools and Techniques*. (2ª ed.). Estados Unidos: Elsevier Ing.

IIRSA. (2010). Recuperado de <http://www.iirsa.com>

Il-Gyo C., Susan, A. y Chi-hyuck, J. (2007). A data mining approach to process optimization without an explicit quality function. *IIE Transactions*. 39, 795-

Instituto Nacional de Estadística de Bolivia (2010): Retraído de: [www.ine.gob.bo](http://www.ine.gob.bo)

Jiwattanakupaisarn, P., Noland R., Graham, D. y Polak, J. (2008). Highway infrastructure and state-level employment: A causal spatial analysis. *Papers in Regional Science*. 88(1), 133-159.

Kamal, M. (2004). Applications of Data Mining Technology in Highway Safety Management. *Paper Presented at the Transpo 2004. Diciembre 6, 2004*. Miami, Estados Unidos.

Kelly, F. Data and Transport. (2006). *Paper Presented at the Faculty of Mathematics, University of Cambridge. Mayo 9, 2006*. Cambridge, Estados Unidos.

Krugman, P. (2002). *Geografía y comercio*. Barcelona: Antoni Bosch Editor.

Kumar, S. (2003). Data mining applications in transportation engineering. *Transport*. 18(5), 216-223.

Montero, L., Codina, E., Barceló, J. y Barceló, P. (2001). A combined methodology for transportation planning assessment. Application to a case study. *Transportation Research Part C: Emerging Technologies*. 9 (3), 213-230.

Neves, J. (2008). Factores de Localización de postos de combustíveis em Fortaleza. *RAC. Revista de Administração Contemporânea*. 175-192.

Ortúzar, J. (1994). *Modelos de demanda de transporte*. Santiago de Chile: Ediciones Universidad Católica.

Ortúzar, J. (2000). *Modelos Económicos de Elección Discreta*. Santiago de Chile: Ediciones Universidad Católica.

Porter, M. (2002). *Competitive Advantage*. (2a Ed.). Buenos Aires: Grupo Patria Cultural.

Shawe, J., De Bie, T. y Cristianini, N. (2006). Data mining, data fusion and information Management. *IEE Proceedings. Intelligent Transport System*. 153(3), 221-229.

Sheng J., Dianhai W. y Hongsheng, Q. (2010). Bayesian Network Method of Speed Estimation from Single-Loop Outputs. *Journal of Transportation Systems Engineering and Information Technology*. 10(1), 54-58.

Shunji, M. (1999). SIG Manual Base. *Revista Sociedad Latinoamericana de Percepción Remota y Sistemas de Información Espacial, SELPER*. 15(1), 8-36.

Stolzer, A y Halford, C. (2007). Data mining methods applied to flight operations quality assurance data: a comparison to standard statistical methods. *Journal of Air Transportation*. 12(1), 6-24.

Sussman, J. y Sgouridis, S. (2004). Regional strategic transportation planning as a CLIOS. *Paper Presented at the ESD Symposium 2004. Marzo 12, 2004*. Massachusetts, Estados Unidos.

Universidad de los Andes, Venezuela. (2006). Los sistemas de información geográfica. *Geoenseñanza*. 11(1), 107-116.

## **BIOGRAFIA**

M.A.N. Emma Vanessa Casas Medina, egresada del Centro de Estudios Superiores del Estado de Sonora, en Comercio Internacional, estudio la maestra en Administración de Negocios en la Universidad del Valle, profesora de tiempo completo en el Centro de Estudios Superiores del Estado de Sonora en la carrera de Comercio Internacional y actualmente estudia el Doctorado en Logística y Dirección de la Cadena de Suministro. Universidad Popular Autónoma del Estado de Puebla. emma.casas@cesues.edu.mx

MSC. Carlos Alberto González Camargo, egresado de la Universidad Santo Tomás, Bogotá, Colombia, en Ingeniería Civil, además estudio Economía en la Universidad Nacional de Colombia, Bogotá, Especialista en Sistemas de Información Geográfica, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, Especialista en Pedagogía y Docencia Universitaria, en la Universidad La Gran Colombia, Bogotá, Colombia, MSC en Ingeniería con énfasis en Transporte, Universidad de los Andes y actualmente se encuentra estudiando el Doctorado en Logística y Dirección de la Cadena de Suministro. Universidad Popular Autónoma del Estado de Puebla, México, y se desempeña como consultor y en la Universidad Jorge Tadeo Lozano de Bogotá, como decano de Ingeniería Industrial y profesor asociado de la Especialización en Logística Comercial Nacional e Internacional.