

RETO DATAJAM CRC - RETO 1

Duvan Leonardo Chavez Buitrago, Angie Jinet Pacheco Cubillos, Brayan David Ruiz.

Universidad Distrital Francisco José de Caldas
Ingeniería en Telemática
Bogotá, 2020

RESUMEN

Para la solución del Reto 1 del DataJam se abordan las problemáticas presentes en la actualidad, para el acceso a los diferentes servicios de telecomunicaciones en zonas rurales, en las cuales las nuevas tecnologías e infraestructura presentes en el país, no han podido llegar hasta aquellos lugares y por ende, no se cuenta con un servicio de calidad o simplemente ni se cuenta con el servicio; se analizarán diferentes Dataset y se mostrarán las variables que pueden llegar a influir y las diferencias de servicios en varias regiones.

ABSTRACT

For the solution of Challenge 1 of the DataJam, the current problems are addressed, for access to different telecommunications services in rural areas, in which the new technologies and infrastructure present in the country have not been able to reach those places. and therefore, there is no quality service or there is simply no service; Different Datasets will be analyzed and the variables that can influence and the differences in services in various regions will be shown.

PALABRAS CLAVE

DataSet, rural, reto, operador, velocidad, latencia, modelo lineal.

KEY WORDS

DataSet, rural, challenge, operator, speed, latency, linear model.

INTRODUCCIÓN

El uso y apropiación de las tecnologías de la información, tiene como variable principal la forma en la que se está conectado a internet, siendo esta como la red más grande del mundo, en la que se puede obtener cualquier tipo de información al instante, estar al día en noticias y hasta dar a conocer un producto o servicio.

Internet ha llegado a gran parte de los hogares y de las empresas, en este aspecto se ha abierto una brecha digital con las

personas que no tienen acceso al mismo, en los cuales la penetración de Internet y las nuevas tecnologías es muy limitada en algunos casos.

No obstante, en el transcurso del tiempo se ha venido extendiendo el acceso a Internet en casi todas las regiones del país, de acuerdo con los esfuerzos realizados por el ministerio de las tecnologías de la información y las comunicaciones.

OBJETIVO GENERAL

Seleccionar datasets que sirvan de insumo para el proceso de análisis de los actuales problemas de acceso a los servicios de telecomunicaciones, mediante el uso de una herramienta de minería de datos.

OBJETIVOS ESPECÍFICOS

- Consultar y clasificar los conjuntos de datos disponibles en las distintas bases de datos, a fin de obtener la información necesaria.
- Analizar e interpretar los datos conforme a tablas y gráficas generadas el programa informático.
- Describir con base a la información obtenida las problemáticas regionales que se han venido presentando frente a los retardos y velocidad media en los servicios móviles.

ALCANCE

En el presente análisis se tratarán algunos aspectos fundamentales referentes a las dificultades de uso y apropiación de las TIC a nivel nacional, se tomará también casos puntuales como lo es el departamento del Chocó.

Es por eso que abordarán temas como el alcance de la conectividad a internet fija, la velocidad y latencia de los servicios de internet móviles.

JUSTIFICACIÓN

En Colombia poco más de 11 millones de personas viven en zonas rurales; zonas a las cuales el acceso a la tecnología es mínimo y la infraestructura para la conexión a diferentes redes es casi inexistente, por lo que el acceso a los diferentes servicios de telecomunicaciones puede llegar a ser muy difícil para estas personas y suelen llegar a existir lugares estratégicos en los que se alcanza este tipo de acceso.

Los retos técnicos y presupuestales son grandes en el contexto de la conectividad en zonas rurales, ya que el llevar hasta allá la infraestructura necesaria, puede traer consigo diferentes problemas tecnológicos, como precios elevados por las distancias

que hay que recorrer y terrenos algo inestables y difíciles de trabajar, por lo que el acceso a internet y conexiones móviles en estos lugares no ha sido implementado en todo el terreno nacional, son muy pocas las zonas a las que ha empezado a llegar este tipo de tecnología y no se ha entrado a ondear en zonas demasiado remotas.

Actualmente el avance en conectividad en Colombia ha sido muy grande, con lo cual la fibra óptica ha llegado a la mayoría de ciudades principales del país y se ha logrado conectar más de 1000 municipios que anteriormente no tenían conexión a internet a esta red de alta velocidad; se han creado diferentes planes y proyectos, para que otros municipios, que por sus condiciones geográficas no han podido lograr esta conexión, en el transcurso del tiempo logren llegar a tener cobertura con diferentes operadores, pero de todas formas siguen quedando por fuera municipios que pueden no llegar a estar contemplados en estos planes.

Los planes nacionales identificaron estos problemas de conectividad, como una necesidad importante para poder llevar la educación a cada hogar, ya que muchas de las escuelas rurales no tienen ni un remota posibilidad de lograr una conexión a internet, por lo que sus estudiantes no tienen ningún contacto con las TIC y los avances que estas han tenido en la actualidad, por lo que la forma de acceder a la educación ha sido detenida en el tiempo y siguen usando métodos de aprendizaje pasado, que aunque pueden lograr importantes resultados, han ido avanzado y se han encontrado diferentes ramas en la virtualidad.

La crisis sanitaria a nivel mundial por la llegada del coronavirus Codiv 19 nos ha mostrado diferentes dificultades y nos ha expuesto a diferentes retos para poder continuar con nuestra vida cotidiana; la mayoría de actividades que realizamos en nuestro diario vivir fueron detenidas y la vida como la conocíamos dio un cambio abrupto para la mayoría de personas, por lo cual se identificó la importancia de los servicios de telecomunicaciones para poder llevar nuestra vida de forma normal y seguir con nuestras actividades tanto de trabajo como de educación, así como de entretenimiento, por lo que las personas que cuentan con los medios económicos y también los medios de hardware para poder hacer uso de estos servicios, han podido sobrellevar de mejor forma estas barreras, pero aquellas personas que no cuentan con los mismos medios y mucho menos con el acceso a la conexión necesaria, han tenido que cambiar de forma drástica su vida. La educación en lugares rurales han sido de los más afectados en esta crisis, ya que si antes era difícil acceder a una educación de calidad, con temas como la virtualidad es cada vez más difícil, por lo que como se mencionó antes, la infraestructura en estos sitios, no es la adecuada y la mayoría de personas no cuentan con los equipos necesarios para el acceso por más que puedan llegar a tener conexión, por lo que suelen tener que hacer uso de diferentes recursos prestados o simplemente desplazarse por largos trayectos para

encontrar tanto un computador, como una red, viéndose de gran medida la brecha de oportunidades en las personas, según sus condiciones económicas y su lugar de vivienda, mostrando claramente cómo estos aspectos si pueden llegar a afectar en gran medida el acceso y la incursión en la innovación y las TIC en nuestro país, por lo que por más que las diferentes personas quisieran incorporarse y hacer uso de estos medios, simplemente puede llegar a ser difícil y casi imposible para algunos, acceder a estos medios.

DESARROLLO

Los datos consultados con base al objetivo establecido fueron tomados de la plataforma de intercambio de datos Postdata (<https://www.postdata.gov.co/>), en la cual se encuentran alojados conjuntos de datos con información estadística de los sectores TIC y Postal. De esta manera se realizó el proceso de búsqueda y selección de datasets que nos brinda información importante para el análisis de las problemáticas regionales relacionadas al acceso de los servicios de telecomunicación.

Con base a lo anterior los datasets seleccionados son:

- Acceso Internet Fijo¹
- Latencia - Servicios móviles²

¹Acceso Internet Fijo | Postdata. (2019). [Conjunto de datos]. Recuperado de <https://www.postdata.gov.co/dataset/suscriptor-es-e-ingresos-de-internet-fijo/resource/7c738136-b1ba-4f2c-9f99-e67b282d6425>

²Latencia - Servicios móviles - Medición de Calidad de Servicios de voz y datos Móviles | Postdata. (2019). [Conjunto de datos]. Recuperado de <https://postdata.gov.co/dataset/medici%C3%B3n-de-calidad-de-servicios-de-voz-y-datos-m%C3%B3viles/resource/4c589a7b-40ea-4c37-bada>

- Velocidad media - Servicios móviles³

La herramienta elegida para el análisis de datos son Microsoft Excel y RapidMiner⁴ bajo una licencia educativa para nuestro uso estudiantil. Esta herramienta permite hacer análisis mediante ‘encadenamiento de operadores’ sobre una aplicación con entorno gráfico que facilita su uso.



Ilustración 1.

La interfaz inicial del programa ofrece barras de herramientas útiles como Repository, Operators, Parameters, un recuadro de ayuda, una barra superior con diferentes vistas que usaremos durante el proceso, etc.

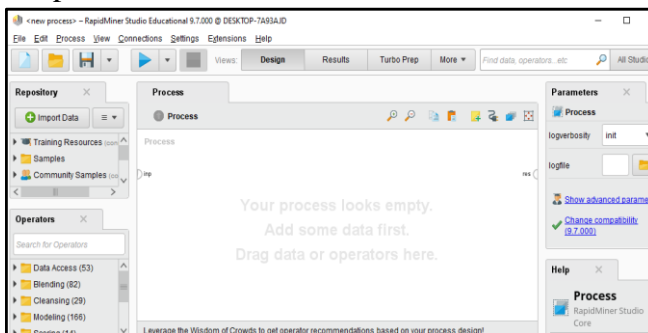


Ilustración 2. Interfaz inicial RapidMiner

Después de realizar la descarga de los archivos CSV correspondientes a los datasets, como primera medida se procedió

³Velocidad media - Servicios móviles - Medición de Calidad de Servicios de voz y datos Móviles | Postdata. (2020). [Conjunto de datos]. Recuperado de <https://postdata.gov.co/dataset/medici%C3%B3n-de-calidad-de-servicios-de-voz-y-datos-m%C3%B3viles/resource/f25ea125-3914-43cd-812b>

a realizar la limpieza de los datos, por ejemplo eliminar caracteres especiales no deseados, pasar todos los datos a mayúsculas con el fin de hacer estadísticas más certeras. Luego se efectuó el cargue respectivo en la herramienta Excel y RapidMiner.

PROCEDIMIENTO

A continuación, se da inicio al procesamiento y análisis de la data seleccionada.

• Acceso Internet Fijo

En una primera medida se puede analizar los distintos tipos de acceso a internet, en este caso se escogió el acceso a internet por servicios Fijos en el territorio nacional.

El dataset proporcionado por postdata brinda información del número de accesos a internet pudiendolas discriminar por departamento y municipio desde al año 2010 hasta el año 2018.

Una primera instancia analizamos la cantidad de acceso por año, como lo muestra la ilustración 3.

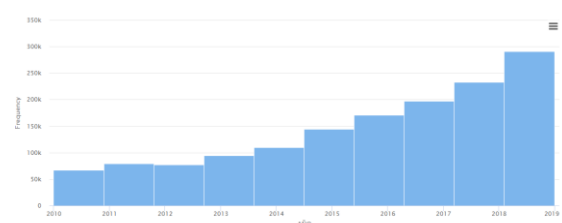


Ilustración 3. Gráfica de Crecimiento Acceso a Internet Fijo en Colombia

⁴ RapidMiner. (2020, 26 febrero). RapidMiner | Best Data Science & Machine Learning Platform. Recuperado 8 de junio de 2020, de <https://rapidminer.com/>

Ilustración 1. [Ilustración]. (2018). RAPIDMINER. Recuperado de <https://www.softwarepublicocolombia.gov.co/es/public-or-civic-software/rapidminer>

El número de accesos ha ido aumentando a media de los años y pues esto evidencia las políticas públicas de aumentar los accesos a internet del ministerio TIC. A nivel nacional como se evidencia en la gráfica el acceso a internet aumentó en 6 veces aproximadamente del año 2010 al año 2018.

Otro aspecto relevante, el uso de internet es fomentado en gran medida a nivel corporativo que, a nivel doméstico, tal como lo muestra la ilustración 4.

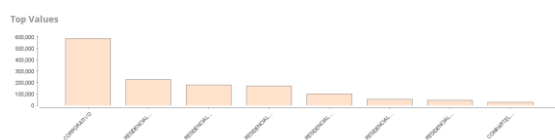


Ilustración 4. Gráfica de Tipo de usuario de Acceso a internet.

Value	Count	Percentage
CORPORATIVO	585,785	40.03%
RESIDENCIAL - ESTRATO 2	228,289	15.60%
RESIDENCIAL - ESTRATO 3	179,887	12.29%
RESIDENCIAL - ESTRATO 1	170,342	11.64%
RESIDENCIAL - ESTRATO 4	102,113	6.98%
RESIDENCIAL - ESTRATO 5	54,425	3.72%
RESIDENCIAL - ESTRATO 6	44,677	3.05%

Ilustración 5. Tabla de Tipo de usuario de Acceso a internet.

El 40.03% del total acumulado de los accesos a internet entre 2010 y 2018 ha sido a nivel corporativo, le siguen los estratos 2 y 3 con 15.6% y 12.29% respectivamente.

Adicionalmente a nivel nacional se tiene un estimado de la tecnología que se usa para acceder a internet, como lo muestra la siguiente ilustración:

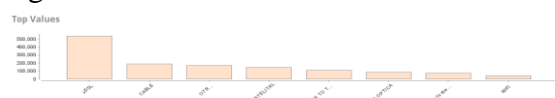


Ilustración 6. Gráfica de Tipo de Tecnología de Acceso a Internet

Siendo la tecnología xDSL la predominante en los últimos 8 años, seguido por cable y satelital.

Value	Count	Percentage
xDSL	532,412	36.39%
CABLE	186,718	12.76%
OTRAS INALAMBRICAS	169,726	11.60%
SATELITAL	144,191	9.85%
FIBER TO THE HOME (FTTH)	107,421	7.34%

Ilustración 6. Tabla de Tipo de Tecnología de Acceso a Internet

A nivel nacional la distribución por departamentos de acuerdo con el número total de conexiones es la siguiente:

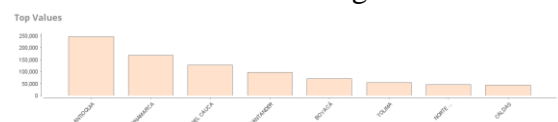


Ilustración 7. Gráfica de Mayores Departamentos con Acceso a Internet

Siendo Antioquia el departamento con más número de conexiones, seguido de Cundinamarca, Valle del cauca y Santander. Estos resultados son esperados, dado que Son los departamentos de las ciudades más grandes del país y están cerca a las respectivas áreas metropolitanas.

Value	Count	Percentage
ANTIOQUIA	246,769	16.87%
CUNDINAMARCA	168,512	11.52%
VALLE DEL CAUCA	125,968	8.68%
SANTANDER	96,717	6.61%
BOYACA	70,650	4.83%
TOLIMA	54,185	3.70%
NORTE DE SANTANDER	45,263	3.09%

Ilustración 8. Tabla de Mayores Departamentos con Acceso a Internet

A nivel de ciudades o municipios, se puede encontrar que las que el mayor número de conexiones se encuentra en las ciudades capitales como lo muestra a continuación.

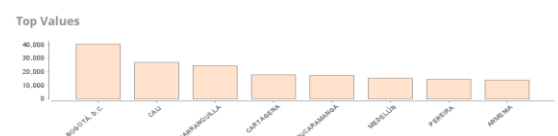


Ilustración 9. Gráfica de Mayores Municipios con Acceso a Internet

Siendo Bogotá, Cali y Barranquilla las ciudades con más número de conexiones a internet a nivel nacional en los últimos años.

Value	Count	Percentage
BOGOTÁ D.C.	40.161	2.74%
CALI	26.717	1.83%
BARRANQUILLA	24.165	1.65%
CARTAGENA	17.558	1.20%
BUCHAMANGA	17.301	1.18%

Ilustración 10. Gráfica de Mayores Municipios con Acceso a Internet

Para un caso puntual de caso de estudio se escogió el departamento de Chocó, donde se realizaron los análisis a continuación. Se obtuvieron datos de la división política Nacional DIVIPOLA del DANE⁵.

Departamento		Municipio		Tipo Municipio / Área no municipalizada	
Código	Nombre	Código	Nombre		
27	CHOCÓ	27001	QUIBDO	Municipio	
27	CHOCÓ	27009	ACACIA	Municipio	
27	CHOCÓ	27025	ALTO BAUDÓ	Municipio	
27	CHOCÓ	27026	ATRAÍTO	Municipio	
27	CHOCÓ	27073	BAUDÓ	Municipio	
27	CHOCÓ	27075	BAJA SOLANO	Municipio	
27	CHOCÓ	27077	BAJO BAUDÓ	Municipio	
27	CHOCÓ	27089	BOJAYÁ	Municipio	
27	CHOCÓ	27135	EL CANTÓN DEL SAN PABLO	Municipio	
27	CHOCÓ	27150	CARMEN DEL DARIÉN	Municipio	
27	CHOCÓ	27160	CORTESÍA	Municipio	
27	CHOCÓ	27209	COMODÓ	Municipio	
27	CHOCÓ	27246	EL CORRIENTE ATRATO	Municipio	
27	CHOCÓ	27250	EL LITORAL DEL SAN JUAN	Municipio	
27	CHOCÓ	27351	ITIMBA	Municipio	
27	CHOCÓ	27372	JANACO	Municipio	

Ilustración 11. Tabla de DIVIOLA del departamento del Chocó.

En los datos obtenidos de Posdata, se encontró que solo existen informes para el el departamento del Chocó entre el 2010 y 2013, en los cuales se analizan a continuación:

Los 5 Municipios con mayor número conexiones a internet dentro del Chocó, se encuentran a continuación:

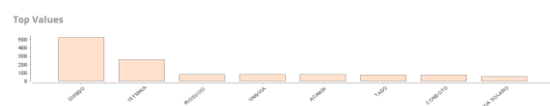


Ilustración 12. Gráfica de Mayores Municipios de Chocó con Acceso a Internet

En donde Quibdó, es la ciudad capital del departamento del Chocó y es el municipio con el 27% de las conexiones a internet, seguido de Istmina y Riosucio con 13,47% y 4.23% respectivamente.

Value	Count	Percentage
QUIBDO	523	27.00%
ISTMINA	261	13.47%
RIOSUCIO	82	4.23%
UNGUIA	82	4.23%
ACANDI	79	4.08%

Ilustración 13. Tabla de Mayores Municipios de Chocó con Acceso a Internet

Por otro lado, los 5 municipios con menor número de conexiones a internet se encuentran a continuación:

Value	Count	Percentage
MEDIO ATRATO	27	1.39%
BAGADO	23	1.19%
JURADO	22	1.14%
EL CANTON DEL SAN PABLO	16	0.83%
CERTEGUI	12	0.62%

Ilustración 14. Tabla Municipios de Chocó con menor Acceso a Internet

También se encuentra que todos los 30 municipios que reporta la DIVIPOLA del DANE, cuentan con conexión a internet fijo, así sea con un número muy limitado, de las mismas.

A nivel del uso de las conexiones fijas a internet, se muestra a continuación:

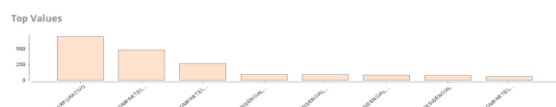


Ilustración 15. Gráfica de tipo de Segmento de Acceso a Internet en el Chocó

Se evidencia que el segmento que mayor número de conexiones sigue siendo el corporativo, telecentro y las instituciones educativas.

Value	Count	Percentage
CORPORATIVO	693	35.78%
COMPARTEL - TELECENTRO	489	25.25%
COMPARTEL - INSTITUCION EDUC...	265	13.68%
RESIDENCIAL - ESTRATO 3	93	4.80%
RESIDENCIAL - ESTRATO 2	89	4.59%

Ilustración 16. Tabla de tipo de Segmento de Acceso a Internet en el Chocó.

⁵ DANE. (2020). Geovisor de Consulta de Codificación de la Divipola. Recuperado de:

A nivel de las tecnologías que permiten la conexión, se muestra a continuación las 5 más usadas:

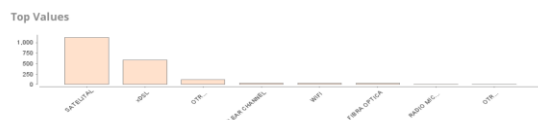


Ilustración 17. Tabla de tipo de Tecnología de Acceso a Internet en el Chocó

Para el caso del Chocó, se evidencia que la tecnología de conexión a internet satelital es la predominante, con la mitad de las conexiones; eso se explica por la dificultad de acceso a los diferentes municipios del departamento.

Value	Count	Percentage
SATELITAL	1,120	57.82%
xDSL	585	30.20%
OTRAS INALAMBRICAS	118	6.09%
CLEAR CHANNEL	39	2.01%
WIFI	33	1.70%

Ilustración 18. Tabla de tipo de Tecnología de Acceso a Internet en el Chocó

También se tomó el dataset: *Internet Fijo Penetración Municipio*⁶ de la plataforma Datos Abiertos. Este brinda información sumariada de Suscripciones a internet Fijo discriminadas por trimestre desde el 2015 hasta el 2019.

Para nuestro caso se tomó los datos desde el 2016 hasta el 2018, en donde los análisis se detallan a continuación.

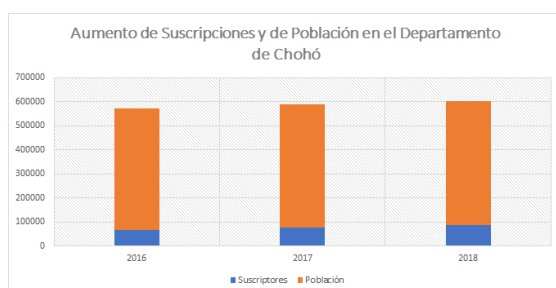


Ilustración 19. Gráfica suscripciones a Internet Fijo y población en el Chocó.

La gráfica muestra número de suscripciones a internet fijo por año y adicionalmente la cantidad de habitantes en todo Chocó. Se evidencia que el aumento es aproximado en la proporción del crecimiento de habitantes de todo el departamento.

De la misma manera se detallan el total de suscripciones en los 7 municipios con más conexiones a internet, tal como lo muestra la ilustración a continuación:

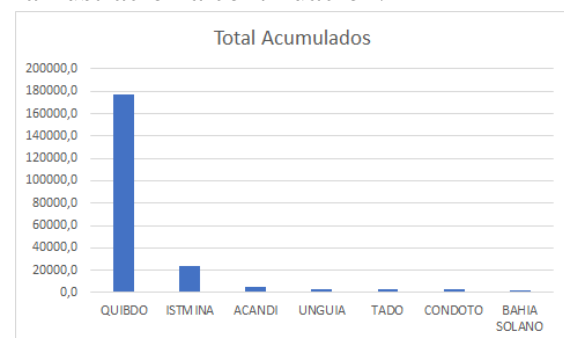


Ilustración 20. Gráfica de Mayores municipios con suscripción a Internet en el Chocó.

Siendo Quibdó, la capital del departamento con la gran mayoría de suscripciones, seguido de Acandí y Unguía.

Finalmente se presenta los 7 municipios con menor de suscripciones a internet fijo en el departamento, como lo muestra la imagen a continuación:

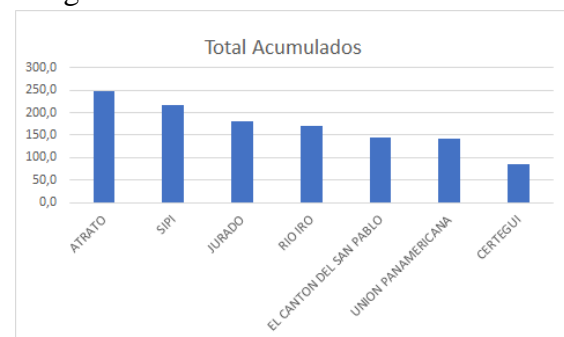


Ilustración 21. Gráfica municipios con menor número de suscripciones a Internet en el Chocó.

⁶ Datos Abiertos. (2020). Internet Fijo Penetración Municipio. Recuperado de: <https://www.datos.gov.co/Ciencia-Tecnolog-a->

[e-Innovaci-n/Internet-Fijo-Penetraci-n-Municipio/fut2-keu8](https://www.datos.gov.co/Ciencia-Tecnolog-a-)

Cabe resaltar que estos valores de suscripción entre 2016 y 2018, han aumentado muy poco desde el 2013, donde son tomados los datos de la anteriores gráficas de posdata. Los municipios con mayor y menor número de conexiones, en los dos datasets son prácticamente los mismos. Únicamente Quibdó refleja un aumento significativo en el número de conexiones fijas a internet.

- **Latencia - Servicios móviles.**

Este recurso disponible en Postdata presenta la experiencia de usuario con respecto al factor de latencia que este pueda presentar al momento de acceder a audio, video y videojuegos desde servicios móviles.

La latencia medida con respecto a la rapidez con la que viajan los datos desde el origen de la solicitud hasta el destino se presenta en una tabla que relaciona el año, ciudad, operador, valor y la descripción; tal como se muestra en la siguiente captura:

	A	B	C	D	E
1	anno	ciudad	operador	valor	KPI
2	2015	Barranquilla	Avantel	201,33	Ping - Latencia (ms)
3	2015	Barranquilla	Claro	210,54	Ping - Latencia (ms)
4	2015	Barranquilla	Etb	165,32	Ping - Latencia (ms)
5	2015	Barranquilla	Éxito	228,56	Ping - Latencia (ms)
6	2015	Barranquilla	Movistar	172,21	Ping - Latencia (ms)
7	2015	Barranquilla	Tigo	153,14	Ping - Latencia (ms)
8	2015	Barranquilla	Virgin	316,43	Ping - Latencia (ms)
9	2015	Bogotá	Avantel	144,4	Ping - Latencia (ms)
10	2015	Bogotá	Claro	206,63	Ping - Latencia (ms)
11	2015	Bogotá	Etb	126,85	Ping - Latencia (ms)
12	2015	Bogotá	Éxito	153,1	Ping - Latencia (ms)
13	2015	Bogotá	Movistar	131,73	Ping - Latencia (ms)
14	2015	Bogotá	Tigo	139,55	Ping - Latencia (ms)
15	2015	Bogotá	Virgin	486,59	Ping - Latencia (ms)
16	2015	Bucaramanga	Avantel	380,32	Ping - Latencia (ms)

Ilustración 22. Tabla de datos de latencia.

El dataset contiene un total de 196 registros produciendo un archivo total de 10.13 KB.

Se importó la data en el RapidMiner usando el carácter separador punto y coma (;). A

partir del cargue aparece una vista con la tabla de datos y se creó un modelo automático.

Con base a las opciones que brinda la herramienta, se generó un modelo lineal para la predicción de la columna *valor*, en donde se relaciona el valor original, la ciudad, el operador y el año. De esta manera en la ilustración 4 se observa las ciudad que con base a la predicción se encuentran en altos valores de latencia como en el caso de Quibdó, lo que generará una experiencia de usuario negativa frente a otras ciudad como Bogotá y Medellín donde la predicción arrojó valores bajos/medios. De igual manera se observa el posicionamiento de los diferentes operadores en cada una de las ciudades en cada periodo de año.

Generalized Linear Model - Predictions

Row No.	valor	prediction(valor)	ciudad	operador	ño año
1	201.330	313.681	Barranquilla	Avantel	2015
2	228.560	302.445	Barranquilla	Éxito	2015
3	172.210	87.763	Barranquilla	Movistar	2015
4	153.140	248.103	Barranquilla	Tigo	2015
5	316.430	163.008	Barranquilla	Virgin	2015
6	144.400	196.025	Bogotá	Avantel	2015
7	206.630	159.076	Bogotá	Claro	2015
8	126.850	187.157	Bogotá	Etb	2015
9	131.730	-29.893	Bogotá	Movistar	2015
10	184.850	357.094	Medellín	Etb	2015
11	299.330	354.726	Medellín	Éxito	2015
12	645.220	215.288	Medellín	Virgin	2015
13	703	752.158	Quibdó	Etb	2015
14	469	749.790	Quibdó	Éxito	2015
15	705	695.448	Quibdó	Tigo	2015
16	132.460	234.300	Bogotá	Avantel	2016
17	140.310	225.432	Bogotá	Etb	2016
18	177.610	8.382	Bogotá	Movistar	2016
19	168.650	337.884	Bucaramanga	Avantel	2016
20	276.490	300.935	Bucaramanga	Claro	2016

Ilustración 23. Tabla de predicción modelo de regresión lineal en la latencia de servicios móviles.

A partir del modelo lineal se genera la gráfica de predicciones donde cada punto representa una predicción con su respectivo valor verdadero.

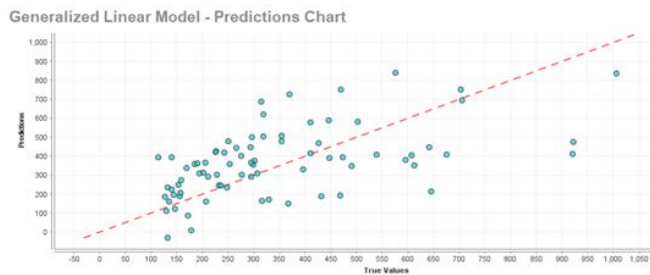


Ilustración 5. Gráfica predicción modelo de regresión lineal - Latencia, servicios móviles.

De forma consecutiva se generó la gráfica de predicción desde el módulo aprendizaje profundo que a partir de un procesamiento diferente al anterior, obtenemos en el plano, puntos en similares ubicaciones.

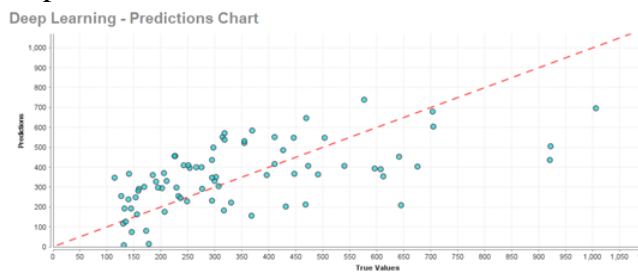


Ilustración 24. Gráfica de predicción aprendizaje profundo - Latencia, servicios móviles.

En ambas gráficas se observan distintos puntos a los largo del plano, unos más cerca de la línea punteada que otros. Esto se puede interpretar en cuanto más cercanos se encuentren los puntos (predicciones) de la línea punteada naranja, mejor será el modelo. De tal forma la observación de los datos se centrará en los puntos que más se encuentren cerca de la línea.

Para obtener datos más acertados y que permita ver en detalle el comportamiento en cada ciudad se realiza la implementación del modelo para generar visualizaciones detalladas como se muestra en la ilustración 7.

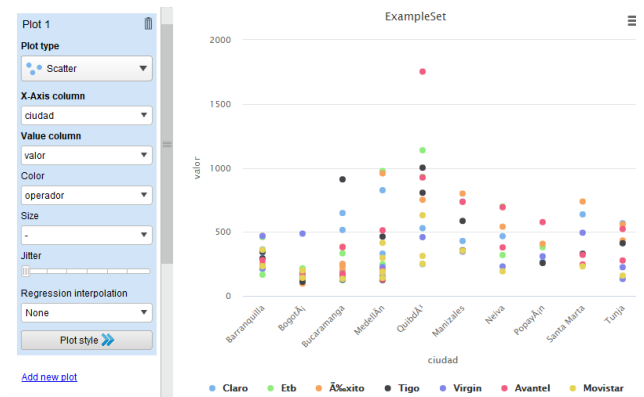


Ilustración 25. Gráfica valor vs ciudad y con referencia a los operadores.

En la construcción de la gráfica anterior se eligieron los parámetros valor y ciudad, además de ubicar por color los operadores disponibles en la data original. El resultado obtenido muestra que las ciudades con menor valor de latencia aparecen con los puntos más agrupados entre sí y más cercanos al eje X. Se logra diferenciar además, a nivel nacional los operadores que presentan mayores niveles de latencia, es decir que no cuentan con buena cobertura en esas zonas geográficas ya sea por falta de infraestructura o por condiciones geográficas.

Las ciudades como Barranquilla, Medellín, Bogotá, Tunja y Popayán a pesar de tener ciertos picos de latencia media con algunos operadores en general cuentan latencia controlada. En el caso de ciudades como Manizales, Bucaramanga y Santa Marta, se presentan numerosos puntos con mediana-alta latencia, pero aún cuentan con operadores que mantienen una buena calidad y baja latencia como en el caso de movistar y claro. Particularmente la ciudad de Quibdó presenta los niveles más altos de latencia; en términos generales la mayoría de los operadores en esa región del país cuentan con bajos indicadores de calidad, lo que se refleja en una pésima experiencia de usuario en el contexto de servicios móviles.

- **Velocidad media - Servicios móviles.**

En este dataset se puede observar la rapidez en la que se descarga y carga contenido desde una página web. De la cual se identifica que, a mayor velocidad, es mejor la experiencia del usuario, y en caso contrario, cuando la velocidad es muy baja, para el usuario la experiencia es negativa.

Los datos obtenidos son 161 y muestran el año en que se tomaron, la ciudad, el operador y la velocidad.

anno	ciudad	operador	velocidad	kpi	entry...
2016	Bogotá	Avantel	18.12	Velocidad Media H...	1
2016	Bogotá	Claro	2.86	Velocidad Media H...	2
2016	Bogotá	Etb	4.73	Velocidad Media H...	3
2016	Bogotá	Éxito	11.96	Velocidad Media H...	4
2016	Bogotá	Movistar	12.72	Velocidad Media H...	5
2016	Bogotá	Tigo	11.69	Velocidad Media H...	6
2016	Bogotá	Virgin	9.40	Velocidad Media H...	7
2016	Bucar...	Avantel	12.41	Velocidad Media H...	8
2016	Bucar...	Claro	3.51	Velocidad Media H...	9
2016	Bucar...	Etb	10.82	Velocidad Media H...	10
2016	Bucar...	Éxito	6.99	Velocidad Media H...	11
2016	Bucar...	Movistar	6.71	Velocidad Media H...	12
2016	Bucar...	Tigo	10.76	Velocidad Media H...	13
2016	Bucar...	Virgin	5.33	Velocidad Media H...	14
2016	Medell...	Avantel	15.63	Velocidad Media H...	15
2016	Medell...	Claro	3.90	Velocidad Media H...	16
2016	Medell...	Etb	6.20	Velocidad Media H...	17
2016	Medell...	Éxito	9.74	Velocidad Media H...	18

Ilustración 26. Tabla de datos velocidad media.

Con la ayuda de RapidMiner, analizamos los datos y realizamos una predicción de estos; en la cual podemos hallar diferentes modelos que nos ayudan a tomar decisiones y predecir velocidades futuras y demás.

En el modelo de regresión lineal, se realizan diferentes cálculos matemáticos, con los cuales se puede aproximar la relación de dependencia entre una variable dependiente y una variable independiente, con lo cual, se puede llegar a comprender qué factores hacen que una variable aumente o disminuya, y en este caso, relacionar la ciudad con la velocidad y también el operador que hace que se logren estas velocidades.

En la siguiente gráfica, se puede observar los valores arrojados por el modelo de regresión lineal, en este se pueden ver las predicciones respecto a la velocidad, y la relación directa que se tiene con la ciudad y el operador; según el color se marcó las ciudades en las que el servicio y la velocidad es más crítica, pero se puede ver que estos valores cambian según el operador en cada ciudad.

Generalized Linear Model - Predictions

Row No.	Velocidad	prediction(Velocidad)	ciudad	operador	year
8	6568	707891213.896	Barranquilla	Virgin	2017
9	12.200	1212285093.235	Bogotá	Avantel	2017
10	10.170	1879139794.157	Bogotá	Éxito	2017
11	7.670	3285003657.957	Bogotá	Virgin	2017
12	11.700	1988100431.951	Bucaramanga	Éxito	2017
13	10.200	2260115836.150	Bucaramanga	Movistar	2017
14	12.030	1808763730.469	Bucaramanga	Tigo	2017
15	10.660	536764618.915	Manizales	Claro	2017
16	4.560	984460640.482	Manizales	Éxito	2017
17	6.760	2390324504.283	Manizales	Virgin	2017
18	12.260	2288734803.581	Medellán	Claro	2017
19	10.210	2557094123.666	Medellán	Tigo	2017
20	7.510	4142294688.949	Medellán	Virgin	2017
21	7.990	984460640.482	Neiva	Éxito	2017
22	13.820	-1291419542.225	Popayán	Claro	2017
23	13.650	-1281576062.295	Popayán	Etb	2017
24	13.130	-571708116.459	Popayán	Movistar	2017
25	12.250	-1023060222.141	Popayán	Tigo	2017
26	8.050	562140343.143	Popayán	Virgin	2017
27	3038	984460640.482	Quibdó	Éxito	2017

Ilustración 26 y 27. Tabla de predicción modelo de regresión lineal.

Con el modelo de regresión lineal, también podemos generar una gráfica, en la cual se identifican las variables y su relación entre ellas, con lo cual mientras más cerca se encuentren a la línea del centro, son mejores los datos analizados para la predicción.

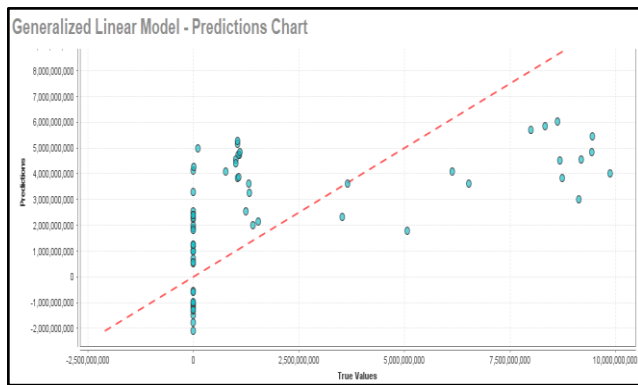


Ilustración 28. Gráfica predicción modelo de regresión lineal.

En la siguiente gráfica, se puede observar un tipo de modelo diferente, este es llamado aprendizaje profundo, en la que se realizan diferentes predicciones y al final se puede observar una gráfica similar a la anterior, en la que los puntos marcados demuestran la importancia de los datos y la dependencia entre sí.

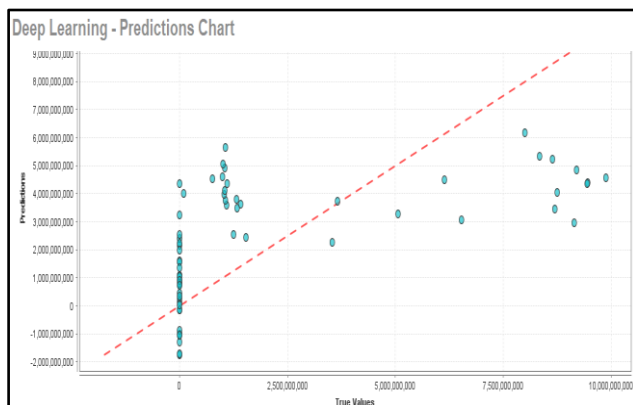


Ilustración 29. Gráfica predicción aprendizaje profundo.

En la siguiente gráfica, se analizaron dos variables las cuales fueron velocidad y ciudad, para así poder identificar cómo puede influir una en la otra, por lo que, según la ciudad y el operador, la velocidad cambia drásticamente, por ejemplo, en la ciudad de Bogotá la velocidad del operador Claro es mucho mejor, que los demás operadores.

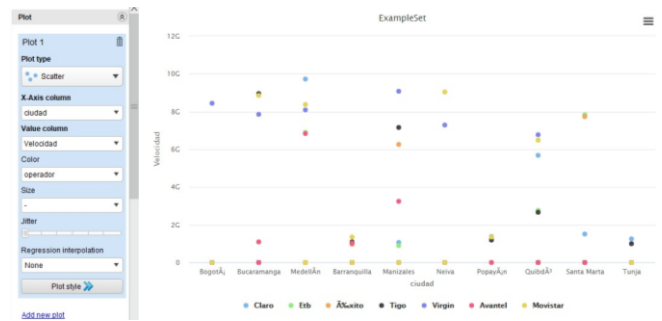


Ilustración 30. Gráfica velocidad vs ciudad.

En la siguiente gráfica se realiza una comparación de la velocidad respecto al operador, en la cual se ve como cada operador cuenta con diferentes velocidades y así mismo llega a las diferentes ciudades con una percepción diferente ante los usuarios, ya que ninguna operador llega a todas las zonas con las mismas velocidades, y estas pueden variar según la cobertura y donde se encuentra la sede principal de cada una de ellas.

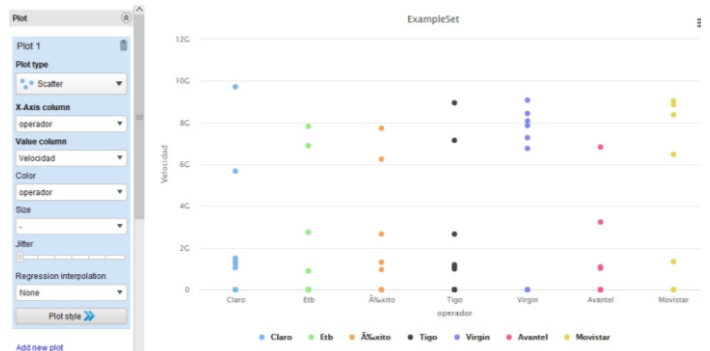


Ilustración 31. Gráfica velocidad vs operador.

CONCLUSIONES

- A nivel de conectividad con de internet fijo a nivel nacional, se ha evidenciado el crecimiento y expansión para conectar al país, pero sin embargo sigue existiendo una brecha digital muy grande en zonas apartadas como lo es el Chocó. Al comparar los datos de la conectividad de 2010 a 2013 suministrados por Posdata y los Datos entre 2016 y 2019

suministrados por Datos Abiertos, el número de conexiones y suscripciones en su mayoría son a nivel corporativo y el crecimiento en el Chocó ha sido muy mínimo en comparación con otras regiones del país, en la misma escala temporal.

Con base al análisis de latencia en servicios móviles tomando diferentes ciudades de Colombia y distintos operadores; se observó de manera clara las diferencias que se han presentado en los últimos años frente a la experiencia de usuario. Se encuentran ciudades con niveles buenos como el caso de Bogotá y Barranquilla donde todos los operadores presentes en esas zonas mantienen una experiencia de usuario para el acceso de audio, video y videojuegos es muy buena; en la ilustración 7 se ven los puntos respectivos a esta ciudad, por debajo del primer límite de valor. Hay otras ciudades con niveles aceptables de latencia entre algunos operadores como Medellín, Tunja, Neiva, entre otros, donde la gran mayoría se encuentran por debajo del primer límite, pero sin olvidar que hay algunos operadores donde la experiencia de usuario es mala.

Finalmente se encuentra el caso específico de Quibdó, donde los valores de latencia son muy altos reflejados en los puntos de la ilustración 7. En general los operadores proveen de insuficiente cobertura en esa zona del país lo que incrementa notablemente los fallos y latencias en el servicio móvil que consumen los habitantes.

De esta manera se logra identificar un problema que se ha venido desarrollando a lo largo de los años referente a la centralización de los recursos del estado y la poca inversión en zonas apartadas del país. En el caso de Quibdó se dispone de infraestructura para proveer servicios móviles de conexión, pero servicios con mucha latencia que genera pésimas experiencias para el usuario; en comparación con otras ciudad, se evidencia que ciudades grandes o económicamente fuertes, poseen mejores estándares de calidad para estos servicios. Por un lado, los operadores privados no invierten lo suficiente en infraestructura y por el otro el Estado hasta hace menos de seis meses, no tenía como prioridad el mejoramiento en cobertura de las zonas más apartadas, hablando en especial de zonas rurales donde por las condiciones geográficas y el difícil acceso, los servicios de conexión no llegan a todas las personas y los lugares donde sí, llegan con señales débiles de 2G.

Se puede decir que para este caso siendo el número de conexiones a internet fijo, la velocidad y latencia en internet móvil variables que afectan directamente el uso y apropiación de las TIC