

Métodos de aprendizaje de maquina para inferir el nivel de cobertura de banda ancha fija en municipios de México

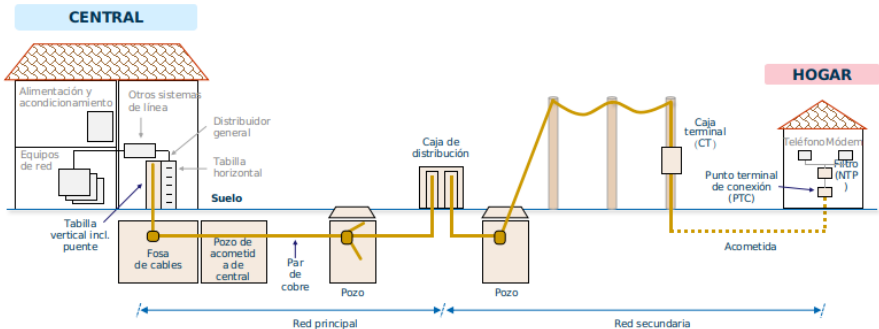
César Zamora Martínez

December 16, 2019

- 1 Cobertura de banda ancha fija en México
- 2 Problemas a explorar

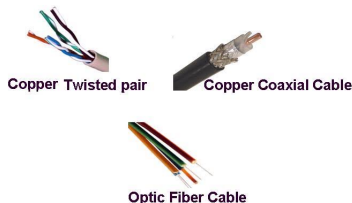
Velocidad y red de acceso

Diagrama ilustrativo de la de red acceso



Velocidad y red de acceso

La velocidad de Internet se encuentra limitada, por la tecnología de acceso.



Interés de este trabajo

Cobertura de cable coaxial y fibra óptica (alta velocidad); requiere grandes inversiones; solo se despliegan en zonas densamente pobladas o con recursos.

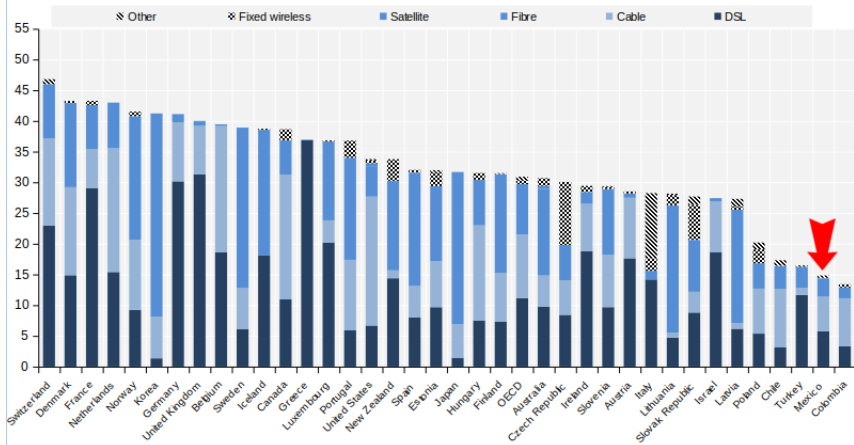
¿Cómo está México en cobertura de banda ancha fija?

- México: 18.9 millones. de accesos + 120 millones de habitantes.
- Accesos: 22% fibra + 37% cable coaxial (total 59%)
- OCDE, proxy de los suscriptores de BAF por cada 100 habitantes la región ¹:

$$\text{Accesos por cada 100 habitantes} = \frac{\text{Accesos}}{\text{Habitantes}} \times 100 \quad (1)$$

¹<http://www.oecd.org/internet/broadband/broadband-faqs.htm>

OECD Fixed broadband subscriptions per 100 inhabitants, by technology, December 2018



¿Cómo está México en cobertura de banda ancha fija?

- En México hay 2,457 municipios,
- ¿cómo se ve la cobertura de fibra óptica y cable coaxial en ellos?
- Veamos el mapa (Junio 2019).
- Cobertura municipal: dependemos de mucha información de operadores que no siempre está disponible (públicamente) o actualizada
- ¿Se puede inferir el nivel de cobertura con información alterna?

Problemas de clasificación de cobertura

A nivel municipal...

P1: ¿Existe o no penetración de BAF de fibra óptica o cable coaxial?

P2: ¿Cuál es el nivel de penetración de BAF de fibra óptica o cable coaxial?

Table: Niveles de penetración en un municipio

| Nivel de penetración | Rango de penetración |
|----------------------|-------------------------------------|
| Muy Alta | $Penetracion > Media\ OCDE$ |
| Alta | $20 < Penetracion \leq Media\ OCDE$ |
| Media | $10 < Penetracion \leq 20$ |
| Baja | $0 < Penetracion \leq 10$ |
| Nula | $Penetracion = 0$ |

Información socio-demográfica de municipios:

- Encuesta Intercensal 2015 (INEGI),
- Información de índice de marginación 2015 (CONAPO)
- Accesos por tecnología, Junio 2019 (IFT)
- Ingreso per cápita "Índice de Desarrollo Humano" (PNUD-ONU)

Pipeline:

- Modelos: Regresión logística, Random Forest, Gradient Tree Bost
- Gridsearch, calibrar los posibles hiper-parámetros de los modelos + validación cruzada,
- Comparar modelos,
- Analizar los resultados

P1 - Variables principales :

- hogares, habitantes
- Densidad hogares y habitantes por km^2
- % hogares sin energía eléctrica
- % pob. en localidades de menos de 5,000 habitantes
- Ingreso anual per cápita (proedi)
- % Habs. que ganan menos de 2 SMM,
- % hogs con tv paga + teléfono fijo/celular

P2 - Variables principales :

Se requieren información adicional para inferir un nivel específico de cobertura a nivel geográfico; (por ejemplo; nivel de competencia en infraestructura/servicios)

- Mismas que en **P1**
- Indicador de si hay 2 o más operadores de BAF
- % hogares con Internet

P1 - Resultados

| # | Modelo | Precisión | Recall | F1-score | Soporte |
|---|------------------------|-----------|--------|----------|---------|
| 1 | Logístico | 0.87 | 0.87 | 0.86 | 612 |
| 2 | Bosque Aleatorio | 0.87 | 0.86 | 0.87* | 612 |
| 3 | Gradient Tree Boosting | 0.87 | 0.86 | 0.86 | 612 |

Table: Promedio ponderado de resultados obtenidos para P1

P1 - Resultados - Matrices de confusión

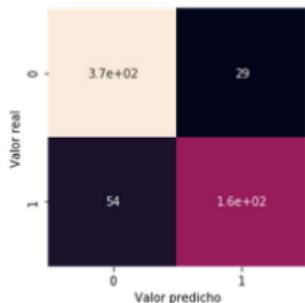


Fig. 10. Matriz de confusión del modelo de bosque aleatorio- P1

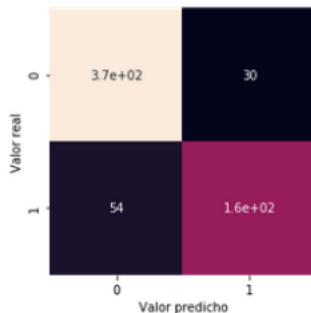
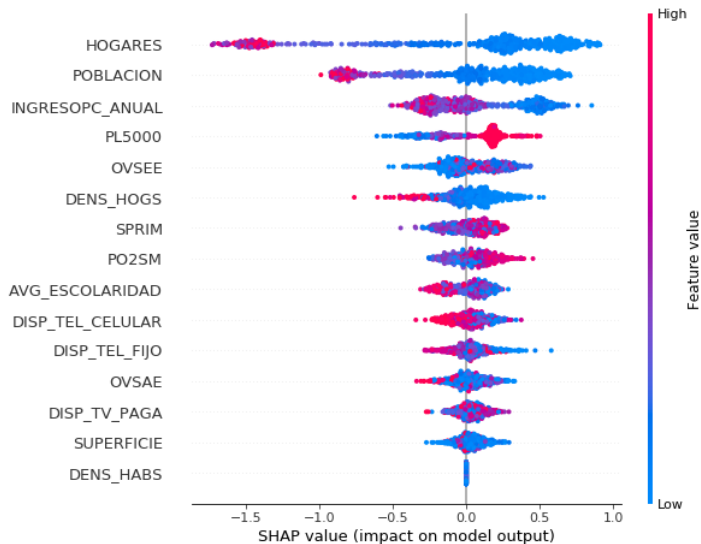


Fig. 11. Matriz de confusión del modelo gradient tree boost- P1

P1 - Resultados - Shap municipios sin cobertura



P2 - Resultados

| # | Modelo | Precisión | Recall | F1-score | Soporte |
|---|------------------------|-----------|--------|----------|---------|
| 1 | Logístico (1 vs all) | 0.90 | 0.88 | 0.89 | 612 |
| 2 | Logístico (all vs all) | 0.90 | 0.88 | 0.89 | 612 |
| 3 | Bosque aleatorio | 0.90 | 0.89 | 0.89* | 612 |
| 4 | Gradient Tree Boosting | 0.89 | 0.88 | 0.88 | 612 |

Table: Promedio ponderado de resultados obtenidos para P2

P2 - Resultados - Matrices de confusión

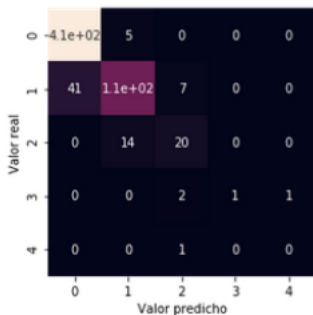


Fig. 16. Matriz de confusión del modelo logístico all vs all - P2

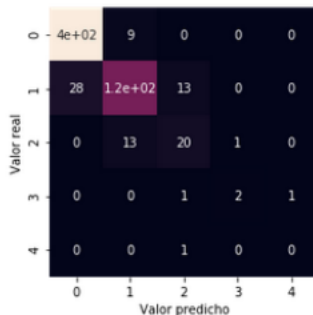
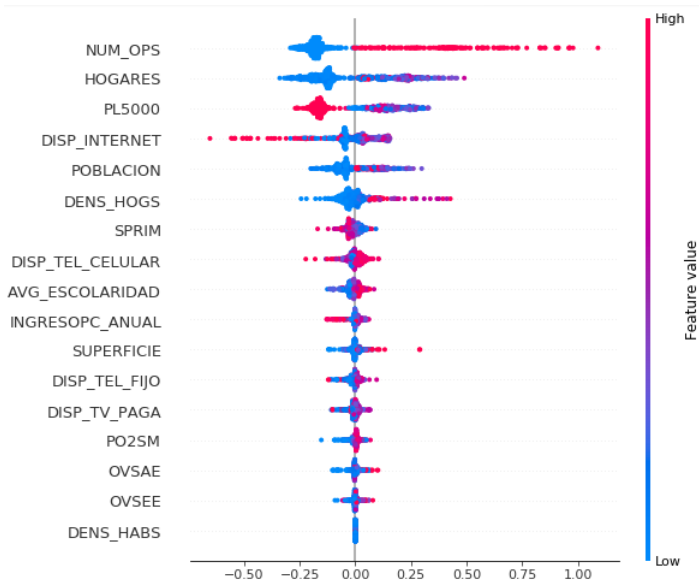


Fig. 17. Matriz de confusión del modelo de bosque aleatorio - P2

Fig. 18

P2 - Resultados - Shap municipios de cobertura Media



Conclusiones y trabajo a futuro

Conclusiones:

- Los métodos de aprendizaje de máquina muestran resultados prometedores para inferir cobertura de BAF ante falta de información.
- La falta de información sobre cobertura puede ser combatida con datos de fuentes públicas (encuestas, estudios económicos, otros)
- Métodos de aprendizaje de máquina pueden ayudar a establecer factores por los cuales hay un cierto nivel de cobertura en una zona del país.

Conclusiones y trabajo a futuro

Líneas que se podrían explorar:

- Incorporar información a nivel geográfico;
- Considerar distancia a infraestructura que permita ampliar la que los operadores tienen en los municipios; por ejemplo cercanía a centrales, distancia a red troncal CFE/Uninet,
- Incorporar información de cobertura de banda ancha con otras tecnologías (banda ancha móvil 4G/5G)
- Extender metodologías/modelos zonas o países donde sin suficiente información del mercado de BAF (e.g. si tienen un nivel de madurez comparable al de México)

Referencias

- **OCDE:**
<http://www.oecd.org/internet/broadband/broadband-faqs.htm>
- **Banco de información de telecomunicaciones (IFT):**
<https://bit.ift.org.mx/BitWebApp/>
- **Documentos del proyecto:**
<https://github.com/czammar/BandaAnchaFija>
- **Mapas interactivos:**
<https://github.com/czammar/BandaAnchaFija/tree/master/Mapas>