



Optimización de precios en Insurtech

Aplicación de metodologías de data science y machine learning para la optimización de precios dentro de una compañía de seguros digital

Jornadas Actuariales Facultad de Ciencias Económicas
Universidad de Buenos Aires
6.11.2020

¿Qué me gustaría compartir con ustedes?

- _ **Problemas tradicionales** de la profesión <> **nuevos enfoques** de resolución
- _ **Experiencia real** de actuarios dentro de una compañía 100% digital
- _ Mostrar que el **rol del actuario** se está volviendo cada vez más **agnóstico** en relación a las industrias con las que interacciona
- _ Cuán importante es **dedicarle tiempo a las nuevas tecnologías,**
metodologías y formas de trabajar

¿Quiénes somos y cómo trabajamos?

- _ Somos **iúnigo** Seguros Argentina una compañía del **Grupo San Cristóbal**
- _ **100% customer centric** y orientada a los negocios de líneas personales
- _ Lanzamos a fines de **2018** y hoy contamos con #150 colaboradores y **#35.000 clientes** de autos y garantía de alquiler

- _ Nuestra organización tiene un **diseño horizontal orientado al agilismo**
- _ Trabajamos con **chapters en conjunto con células** multidisciplinarias
- _ La práctica de **pricing** es responsabilidad del equipo actuarial, pilar fundamental del equipo de **Data Science**

DATA SCIENCE → **actuarial** + business intelligence + machine learning

¿Cuál es el problema a atacar?

Precio final para un seguro de automotor de un prospecto determinado

¿Qué buscamos?

- _ optimizar la relación customer-lifetime-value y la conversión
- _ supervisar la relación ventas <> sanidad del portafolio
- _ detectar segmentos más rentables
- _ otorgar descuentos de manera inteligente
- _ dirigir campañas de marketing hacia esos segmentos

¿Con qué desafíos nos encontramos?

- _ implementación de cálculos real-time + distintos data sources
- _ feature engineering + orquestación de múltiples modelos predictivos
- _ implementación en el producto digital + cumplimiento del marco regulatorio

¿Qué herramientas utilizamos?

- _ machine learning supervisado+no-supervisado para modelos predictivos y su descomposición
- _ herramientas open source para análisis, desarrollo, pipelines e implementación

¿Cómo calculamos el customer-lifetime-value?

$$\sum_{t=0}^{\infty} [Prima_t * (1 - \text{costos}\%) - Siniestros_t - Asistencias_t] * p_t * (1 + i)^{-t}$$

_ **PRIMA** → prima de tarifa impresa en la póliza o en la cotización **GLM**



_ **SINIESTROS** → prima pura esperada **XGBOOST**



_ **ASISTENCIAS** → costo esperado de asistencias **XGBOOST**



_ **SURVIVAL** → probabilidad de supervivencia **KM - WEIBULL**

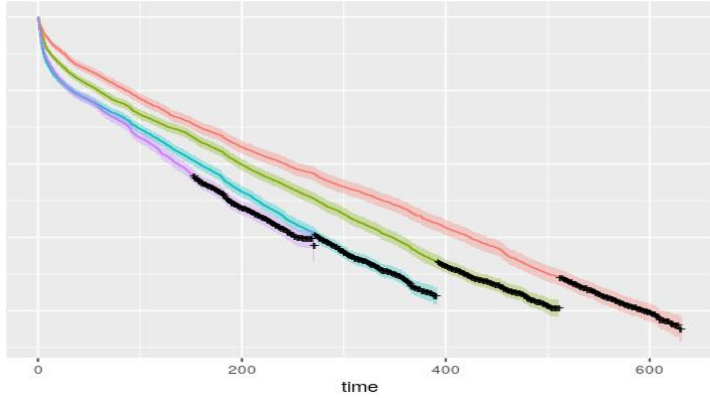


_ **costos%** → otros costos representados como % sobre prima de tarifa

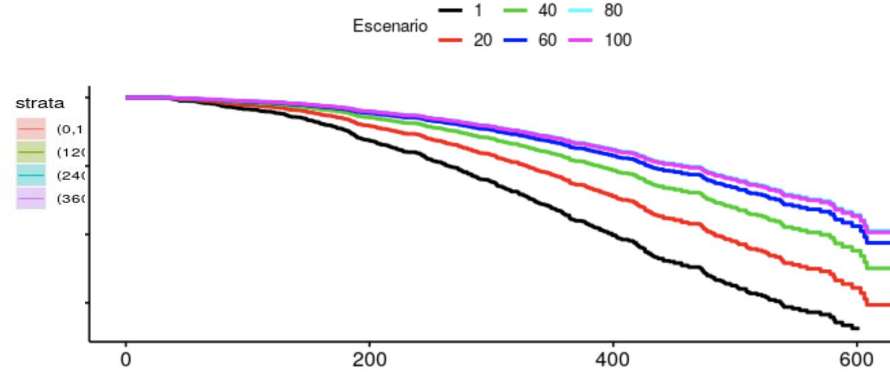
¿Cómo calculamos el customer-lifetime-value?



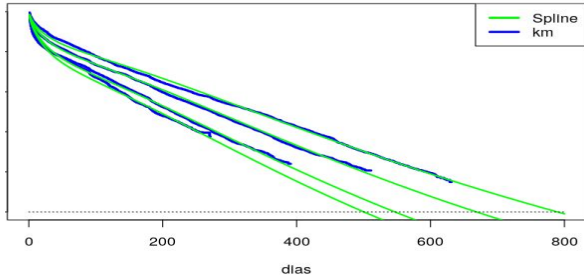
Kaplan-Meier por cohorte (días desde 01-01-2019)



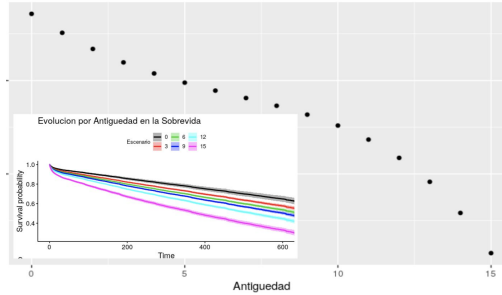
Evolucion por Claim CSAT de la Sobrevida



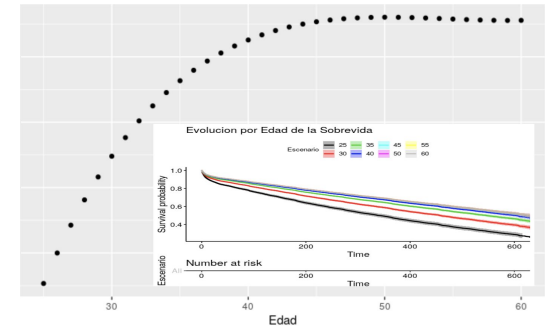
Duracion por cohorte



Evolucion por Antigüedad de la Probabilidad de Supervivencia a 365 días



Evolucion por Edad de la Probabilidad de Supervivencia a 365 días



¿Cómo calculamos el customer-lifetime-value?

POST `https://api.iunigo.com/lifetimevalue` **api que entrega el cálculo** Send Save

Params Authorization Headers (11) **Body** Pre-request Script Tests Settings Cookies Code

none form-data x-www-form-urlencoded raw binary GraphQL JSON Beautify

```
1 {
2   "age": "40",
3   "car_age": "2015",
4   "zone": "11",
5   "zero_km": "true",
6   "excess": "30000",
7   "gnc": "false",
8   "car_id": "460907",
9   "stated_amount": 600000,
10  "vehicle_imp": "NO",
11  "vehicle_type": "SED",
12  "vehicle_fuel_type": "NAF"
13 }
```

parámetros del prospecto a evaluar



Body Cookies Headers (9) Test Results Status: 200 OK Time: 424 ms Size: 869 B Save Response

Pretty Raw Preview Visualize JSON

```
1 {
2   "life_time_value": [
3     {
4       "name": "P1",
5       "value": 927.12794748
6     },
7     {
8       "name": "P2",
9       "value": 3075.168941350002
10    },
11    {
12      "name": "P3",
13      "value": 8577.832424400007
14    }
15  ]
16 }
```

manejo de errores + latencia + tamaño de la respuesta

CLTV esperado del prospecto para los distintos productos en el catálogo

¿Cómo supervisamos las ventas?

Conversion rate = ventas/prospectos

Desde lo analítico

- _ Distance to Market → $DTM = \text{Precio}(i\acute{u}nigo) / \text{Precio}(\text{mercado})$
- _ Comparación diaria con miles de cotizaciones benchmark
- _ Determinación de niveles de competitividad en los distintos segmentos
- _ Curva de demanda y análisis de impacto de cambios tarifarios

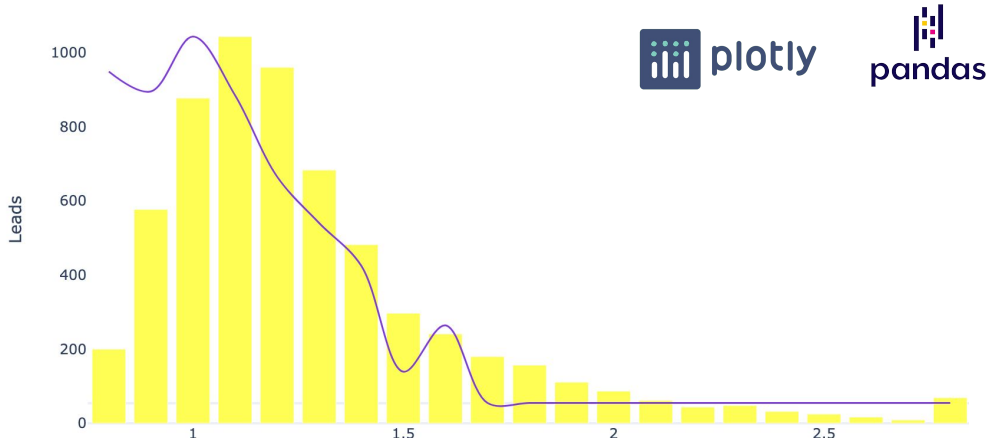
Desde lo predictivo

- _ Modelo de conversión → $y = \text{ventas/prospectos}$
- _ Performance de entrenamiento y baja latencia de respuesta
- _ Variables demográficas + behavioural + feature engineering
- _ RevEng Models de los datos para agregar DTM como feature al modelo de CVR



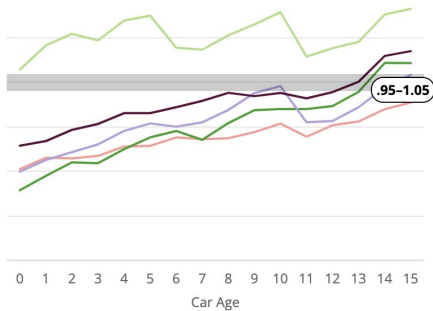
¿Cómo supervisamos las ventas?

Ratio de conversión por rango de DTM



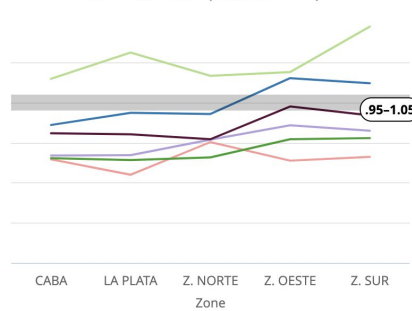
DTM por compañía y antigüedad

DTM: P4 x Antigüedad (últimos 7D)

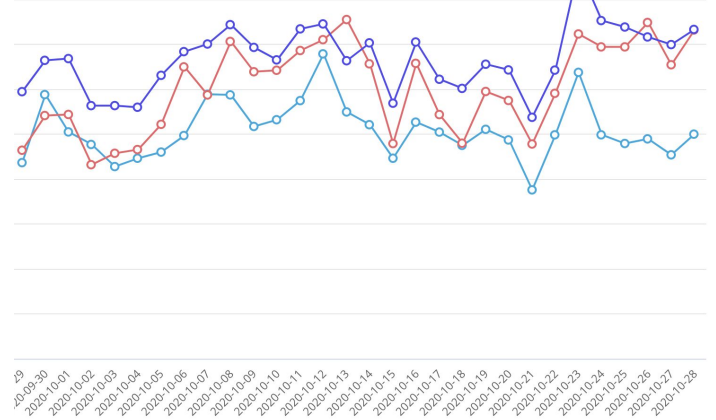


DTM por compañía y zona

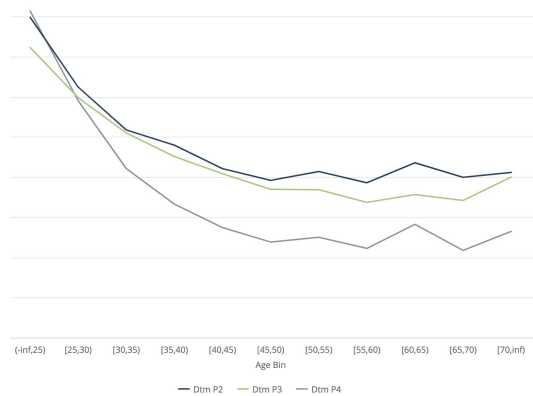
DTM: P4 x Zona (últimos 7D)



Evolución DTM por producto



DTM por producto y edad



Y con todo esto... ¿qué es lo que hacemos?

Identificamos **clusters de CLTV+CVR** y buscamos la **estrategia que maximiza ambas métricas** otorgando **descuentos/recargos sobre el precio base**

Predicción CLTV

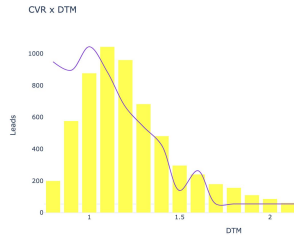


- _ Cálculo de cltv para una BBDD de prospectos
- _ Aplicando algoritmo con el stack de modelos predictivos
- _ precio & siniestros & asistencias & survival

Predicción CVR



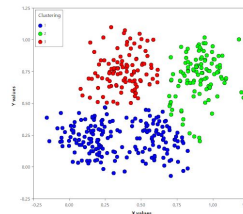
- _ Predicción de la probabilidad esperada de compra para la BBDD de prospectos
- _ Variables demográficas + comportamiento + dtm



Clustering



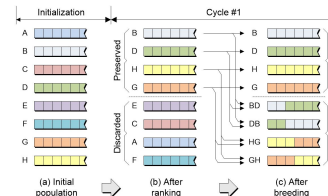
- _ Aplicación de modelo no-supervisado de cluster analysis para detectar segmentos de la combinación cltv+cvr que sean significativamente distintos



Algoritmo Optimización



- _ Aplicación de un algoritmo genético para encontrar la estrategia óptima de descuentos/recargos para los clusters definidos



Output Final



- _ El output final consiste en un nuevo rating factor a incluir en los modelos glm de cotización

| cluster | factor |
|---------|--------|
| 1 | 1.05 |
| 2 | 1.02 |
| 3 | 0.95 |
| 4 | 0.85 |
| 5 | 1.00 |
| ... | |
| 10 | 0.92 |

$$\sum_{i=1}^n [Prime_i \cdot (1 - costos\%) - Siniestros_i - Asistencias_i] \cdot p_i \cdot (1+i)^{-i}$$

¿Cómo automatizamos los pipelines?



~~#10 Pricing~~

#22 Siniestros

#1 Supervivencia

#1 Asistencias

#1 CVR

#7 Rev Eng competencia

#1 Clustering

#1 CLTV

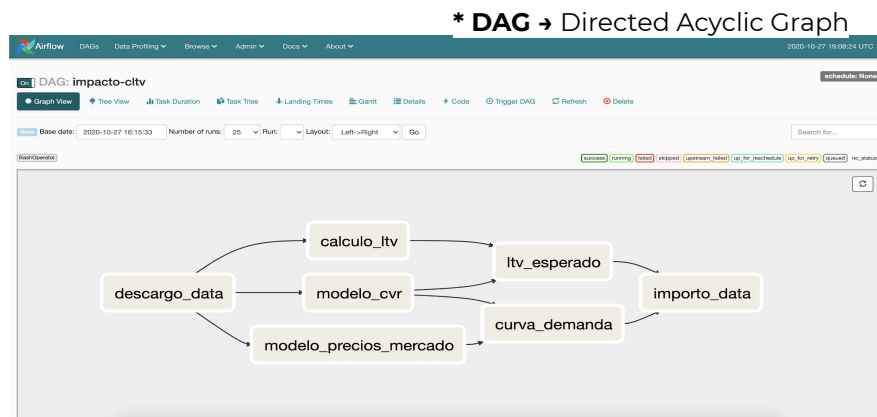
#1 Optimización

~35 pipelines de generación de datos y entrenamiento de modelos

The screenshot shows the Apache Airflow web interface. At the top, there's a navigation bar with 'Airflow', 'DAGs', 'Security', 'Browse', 'Admin', 'Docs', and 'About'. Below the navigation bar, a yellow warning message states: 'The scheduler does not appear to be running. Last heartbeat was received 2020-03-31, 11:34:26. The DAGs list may not update, and new tasks will not be scheduled.' Below the warning, there's a 'DAGs' section with a search bar and a table of DAGs. The table has columns for 'DAG', 'Schedule', 'Owner', 'Recent Tasks', 'Last Run', 'DAG Runs', and 'Links'. The first row shows 'example_dag' with a schedule of '0-05:00' and a last run of '2018-01-01, 00:00:00'. The table shows 10 entries, with the last one being 'example_dag9'.

| DAG | Schedule | Owner | Recent Tasks | Last Run | DAG Runs | Links |
|--------------|----------|---------|--------------|----------------------|----------|-------|
| example_dag | 0-05:00 | airflow | | 2018-01-01, 00:00:00 | | |
| example_dag1 | @daily | airflow | | 2018-01-01, 00:00:00 | | |
| example_dag2 | @daily | airflow | | | | |
| example_dag3 | @daily | airflow | | | | |
| example_dag4 | @daily | airflow | | | | |
| example_dag5 | @daily | airflow | | | | |
| example_dag6 | @daily | airflow | | | | |
| example_dag7 | @daily | airflow | | | | |
| example_dag8 | @daily | airflow | | | | |
| example_dag9 | @daily | airflow | | 2018-01-03, 00:00:00 | | |

Consola Apache Airflow



Diseño del DAG* de un proceso

¿Cómo luce el proceso para el usuario?

iúnigo

El mejor seguro para tu auto está acá.

No busques más, disfrutá nuestro plan más vendido con un 20% off por 3 meses.



Buscá tu auto

Seleccioná la marca de tu auto

¿Cuántas puertas tiene tu auto?

4 Puertas 5 Puertas

¿En qué zona vivís?

CAPITAL FEDERAL

BUENOS AIRES

ROSARIO

CORDOBA

LA PLATA

OTRA

¿Cuántos años tenés?

variables demográficas

variables de comportamiento

1. Cálculo pricing base
2. Cálculo **CLTV** inicial
3. Asignación del **cluster**
4. **Descuento/Recargo** del Cluster

Precio final optimizado

¡Elegí tu plan!

Fecha de inicio: 28/10/2020 Guardar Cotización

FORD FIESTA / Suma aseguradora: \$820000

| | | |
|-----------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| Base: Responsabilidad civil \$3030/mes <small>Te cubrimos robo, incendio y destrucción totales.</small> | Más: Terceros completo \$20% OFF por 3 meses! \$3.826/mes \$4.783 <small>Además te cubrimos: granitos, cristales, camuflado y ruedas sin límites.</small> | Full: Todo riesgo \$6619/mes <small>Te cubrimos los daños asociados: conductor o a terceros que elijas.</small> |
| <input type="button" value="Cotizar online"/> | <input type="button" value="Cotizar online"/> | <input type="button" value="Cotizar online"/> |



iúnigo

Resultados + aprendizajes

- _ Resultados: logramos **CLTV 8x CVR +7.8%**
- _ **Orquestar múltiples pipelines** de datos y entrenamiento de modelos predictivos
- _ Implementar un **proceso complejo con baja latencia**
- _ **Descomponer fenómenos** de conversión (cvr) y valor económico de los clientes (cltv)

Problemas tradicionales de la profesión <> **nuevos enfoques** de resolución
Experiencia real de actuarios dentro de una compañía 100% digital
Mostrar que el **rol del actuario** se está volviendo cada vez más **agnóstico** en
relación a las industrias con las que interacciona → **la versatilidad es clave**
Cuán importante es **dedicarle tiempo a las nuevas tecnologías, metodologías**
y formas de trabajar

Resumen del stack tecnológico

Nuestro **stack** de basa en herramientas y frameworks **open source**

_ *Scripting and prototyping:*   

_ *Data analysis:*  plotly  pandas

_ *ML Model and evaluation:*  LightGBM    SHAP  statsmodels  TensorFlow  LIFELINES

_ *Pipeline orchestrator and scheduler:*  Apache Airflow

_ *API Development:*  Flask  gunicorn

¡Muchas gracias!

Mariano Costa - Chief Data Science Officer



mariano.costa@iunigo.com



mariano-costa