

## REDICTION TASK



What is the type of task?  
Which entity are predictions made on?  
What are the possible outcomes to predict?  
When are outcomes observed?

**Tipo de tarea:** Clasificación binaria.

**Predicción:** ¿Un cliente comprará (1) o no comprará (0) una póliza adicional para un Caravan?

**Resultados esperados:** Probabilidad de compra.

**Momento:** Resultado después de la campaña de contacto con clientes.

**Summary:** Tarea de clasificación binaria servirá para predecir si un cliente contratará (1) o no (0) una póliza de Caravan. El modelo estimará la probabilidad de compra tras una campaña de marketing.

## DECISIONS



How are predictions turned into actionable recommendations or decisions for the end-user? (Mention parameters of the process / application for this.)

**Predicciones:** Las predicciones apoyarán para:

- Soportar las decisiones de segmentación del mercado.
- Priorizar grupos objetivo de las estrategias de marketing.
- Optimizar costos para las estrategias de mercado.
- Buscar un mayor ROI en las estrategias masivas de marcado.

**Métricas:**

- *Tasa de conversión:* Potencial cliente vs contratación de póliza.
- *Efectividad de campañas:* Costo de campañas vs Contratación de pólizas.

## VALUE PROPOSITION



Who is the end beneficiary, and what specific pain points are addressed?  
How will the ML solution integrate with their workflow, and through which user interfaces?

**Beneficiario:** Equipo de marketing y ventas de la aseguradora.

**Problema que resuelve:**

- Mejorar retorno de inversión.
- Optimizar gasto de campañas masivas.
- Aumentar conversión candidato vs cliente.
- Propuesta de valor para integrar al flujo de marketing.

**Integración:** Idealmente deberá integrarse a algún CRM o Herramientas que el equipo de marketing use para difusión masiva: ej. Mailing Segmentado.

## DATA COLLECTION



How is the initial set of entities and outcomes sourced (e.g., database extracts, API pulls, manual labeling)?  
What strategies are in place to update data continuously while controlling cost and maintaining freshness?

**Fuente:** Dataset | BD de la aseguradora holandesa - Sentient Machine Research.

**Obtención:** Registros históricos internos de la compañía.

**Estructura:**

- Col 1 a 43: Datos Sociodemograficos (Relacionados al Código Postal)
- Col 44 a 85: Datos asociados a los productos de seguros (Relacionados a las pólizas existentes contratadas)
- Col 86: Variable binaria objetivo para predicción de contratación para pólizas Caravan (AutoRodante)

**Estrategias ingesta futura:** Actualización periódica con nuevos clientes y resultados de campañas de marketing.

## DATA SOURCES



Where can we get data on entities and observed outcomes? (Mention internal and external database tables or API methods.)

**Entidades:**

- Datos Sociodemográficos (TBD)
- Datos Productos asociados (TBD)
- Predicción de conversión

**Outcomes:** Compra/no compra de póliza.

**Fuentes:**

- *DataSet* interno con datos consolidados. (Fase 1)
- *Base de datos* interna de aseguradora (CRM, ERP). (Fase 2)

**Formato:** Archivos CSV | BD relacionales, ambos con datos históricos.

## IMPACT SIMULATION



What are the cost/gain values for (in)correct decisions?  
Which data is used to simulate pre-deployment impact?  
What are the criteria for deployment?  
Are there fairness constraints?

**Impacto:**

- Valores de predicciones:
  - **Falsos positivos:** Costo por contactar clientes sin interés.
  - **Falsos negativos:** Pérdida de oportunidad de venta (Clientes potenciales).

**Simulación:**

- Experimentar con campañas de tipo Test A/B (Dos grupos MK Tradicional vs ML Predictivo)

## MAKING PREDICTIONS



Are predictions made in batch or in real time?  
How frequently?  
How much time is available for this (including featurization and decisions)?  
Which computational resources are used?

**Tipo:** Batch.

**Razón:**

- Análisis de resultados de las campañas.
- Input para próximas Campañas publicitarias.

**Frecuencia:** Trimestralmente o cada vez que se planea una acción de marketing masiva.

## BUILDING MODELS



How many models are needed in production?  
When should they be updated?  
How much time is available for this (including featurization and analysis)?  
Which computation resources are used?

**Modelos:**

- *Regresión Logística* (Generación de un baseline).
- *Random Forest* (Propuesta más robusta ante ruido, captura interacciones).
- *Gradient Boosting* (XGBoost/LightGBM) (Para mayor precisión)

**Actualización:**

- Trimestral. (4 Actualizaciones al año)
- Bajo demanda dada la importancia de las campañas de marketing y su efectividad.

## FEATURES




What representations are used for entities at prediction time?  
What aggregations or transformations are applied to raw data sources?

**Representaciones:** Datos Categóricos y Numéricos (Evaluar si contamos con edad, ingresos, historial de seguros, transacciones). (TBD)

**Transformaciones:**

- *One-hot encoding* para variables categóricas.
- *Normalización/escalado* en variables numéricas.

**Agregaciones:** Productos contratados, Poder adquisitivo, Tipo de Cliente, Tiempo

<p><b>Sesgos:</b> Riesgo de discriminar ciertos grupos por el uso de Código postal.</p>	<p><b>Recursos:</b></p> <ul style="list-style-type: none"> <li>- Hardware: CPU / Memoria para procesamiento masivo y en paralelo (Explorar recursos en cloud) y Storage para Bases de datos o Datasets.</li> <li>-Software: Librerías Python, Java, Sw libre y propietario.</li> </ul>		<p><b>Recursos:</b></p> <ul style="list-style-type: none"> <li>- Hardware: CPU / Memoria para procesamiento masivo y en paralelo (Explorar recursos en cloud) y Storage para Bases de datos o Datasets.</li> <li>-Software: Librerías Python, Java, Sw libre y propietario.</li> </ul>	<p>como Cliente, Métricas de gasto.</p>
	<div> <div> <p><b>MONITORING</b></p> <p>Which metrics and KPIs are used to track the ML solution's impact once deployed, both for end-users and for the business?</p> <p>How often should they be reviewed?</p> </div> <div> <p><b>Métricas de predicción:</b></p> <ul style="list-style-type: none"> <li>- F1-score, AUC (Área Bajo la Curva)</li> <li>- Precisión y recall.</li> </ul> <p><b>Métricas de negocio:</b></p> <ul style="list-style-type: none"> <li>- Tasa de conversión de campañas.</li> <li>- ROI.</li> </ul> <p><b>Frecuencia de revisión:</b> Trimestral y tras cada campaña de marketing.</p> <p><b>Alertas:</b> Desempeño de los modelos definiendo un umbral aceptable (ej. AUC &lt; 0.70).</p> </div> <div>  </div> </div>			



Version 1.2. Created by Louis Dorard, Ph.D. Licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).  
Please keep this mention and the link to [ownml.co](https://ownml.co) when sharing.

**[OWNML.CO](https://ownml.co)**