

## TABLA DE CONTENIDO

1. Introducción .....	1
1.1. Características de la empresa .....	1
1.2. Descripción del mercado en el que participa la empresa .....	2
1.3. Descripción del servicio en el que se enfoca el trabajo .....	4
2. Problema detectado y justificación del problema.....	6
2.1. Problema existente en el servicio entregado .....	6
2.2. Elección del CPC como métrica de desempeño .....	8
3. Objetivos del trabajo de memoria .....	12
3.1. Objetivo general.....	12
3.2. Resultados esperados .....	13
3.3. Alcances del trabajo de título .....	14
4. Marco conceptual .....	15
4.1. Descripción de decisiones, métricas de desempeño y plataformas relevantes para el proceso de publicación de anuncios digitales .....	15
4.2. Consideraciones generales del proceso de publicación de anuncios en plataformas digitales .....	19
4.3. Importancia de los modelos de atribución para medir efectividad de una publicación digital .....	24
4.4. Marco teórico asociado a algoritmos de aprendizaje automático .....	29
4.4.1. Decision Tree Classifier .....	29
4.4.2. Redes Neuronales.....	31
4.4.3. Random Forest Classifier .....	33
4.4.4. XGBoost Classifier .....	33
4.4.5. Decision Tree Regressor .....	37
4.4.6. Random Forest Regressor.....	37

4.4.7. XGBoost Regressor .....	38
4.5. Métricas de evaluación de desempeño de modelos predictivos .....	38
4.5.1. Accuracy .....	38
4.5.2. Curva ROC (AUC) .....	39
4.5.3. MAE (Mean Squared Error) .....	41
4.5.4. RMSE (Root Mean Squared Error) .....	42
5. Metodología .....	42
5.1. Metodología general utilizada: CRISP-DM .....	42
5.2. Metodología en detalle utilizada para el desarrollo de modelos de aprendizaje automático.....	44
5.2.1. Metodologías generales para todos los algoritmos .....	45
5.2.2. Metodología particular para el desarrollo de modelos de redes neuronales .....	46
5.2.3. Herramientas y tecnologías utilizadas para el desarrollo de los modelos .....	48
6. Entendimiento de los datos.....	49
6.1. Descripción de los datos disponibles .....	49
6.2. Análisis descriptivo de datos .....	53
7. Desarrollo y evaluación de modelos predictivos .....	57
7.1. Modelos de clasificación .....	57
7.1.1. Primer modelo: clasificación de CPC por medio de un umbral único para todas las plataformas .....	58
7.1.2. Segundo modelo: clasificación del CPC por medio de un umbral diferenciado para cada una de las plataformas .....	63
7.1.3. Análisis de importancia de features .....	68
7.1.4. Conclusiones generales de los modelos de clasificación desarrollados .....	69
7.2. Modelos de predicción del CPC como variable continua.....	71

7.2.1. Predicción continua del CPC considerando todas las plataformas al mismo tiempo .....	71
7.2.2. Predicción continua del CPC diferenciando por plataforma .....	72
7.2.3. Conclusiones de los modelos predictivos del CPC como variable continua .....	74
8. Contraste de efectividad de modelos contra decisiones tomadas a criterio experto .....	74
9. Recomendaciones de implementación y propuestas de mejora para trabajos futuros.....	76
9.1. Recomendaciones de uso de modelos para realizar predicciones.....	77
9.2. Recomendaciones relacionadas a la disponibilidad de datos .....	78
9.2.1. Implementación de proceso de registro de datos y consideración de nuevas variables de control .....	78
9.2.2. Consideración de un mayor universo de datos .....	81
9.3. Propuestas para trabajos futuros.....	82
10. Bibliografía.....	84
11. Anexos .....	86
11.1. Resultados predictivos de modelos de clasificación que consideran un umbral de predicción diferente por plataforma .....	86
11.2. Resultados en detalle de modelos de predicción del CPC como variable continua .....	88
11.3. Hiperparametros de cada algoritmo optimizados a través de búsqueda de grilla (sin considerar redes neuronales).....	91
11.4. Procedimiento de optimización de hiperparametros para el algoritmo de redes neuronales .....	93
11.5. Situación de ejemplo expuesta en cuestionario para expertos.....	94
11.6. Ejemplos de publicaciones para cada una de las plataformas.....	95
11.7. Resultado de Shapely Values para modelos XGBoost Classifier (umbral diferenciado).....	99