

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Sobre la empresa . . . . .	1
1.2. Motivación . . . . .	2
1.3. Objetivos . . . . .	3
1.3.1. Objetivo general . . . . .	3
1.3.2. Objetivos específicos . . . . .	3
<b>2. Estado del arte</b>	<b>4</b>
2.1. Fuga en negocios no contractuales . . . . .	4
2.1.1. Construcción del <i>dataset</i> . . . . .	5
2.1.2. Modelos predictivos . . . . .	8
<b>3. Marco teórico</b>	<b>10</b>
3.1. <i>Weight of evidence</i> . . . . .	10
3.2. Reducción de dimensionalidad . . . . .	12
3.2.1. <i>Principal Component Analysis</i> . . . . .	12
3.3. Modelos Estadísticos . . . . .	14
3.3.1. Regresión logística . . . . .	14
3.3.2. <i>Support vector machine</i> . . . . .	18
3.3.3. <i>Decision trees y random forest</i> . . . . .	23
3.4. Metodologías de proyectos . . . . .	27
3.4.1. CRISP-DM . . . . .	27

<b>4. Metodología</b>	<b>29</b>
4.1. Metodología del trabajo . . . . .	29
4.2. Cronología del trabajo . . . . .	32
<b>5. Desarrollo</b>	<b>33</b>
5.1. Entendimiento del negocio . . . . .	33
5.2. Construcción del <i>dataset</i> . . . . .	34
5.2.1. Tablones a utilizar . . . . .	35
5.2.2. Estructuración de la base de datos . . . . .	36
5.2.3. Descripción y construcción de variables independientes . . . . .	37
5.2.4. Segmentación y filtros realizados . . . . .	41
5.2.5. Descripción y construcción de la variable respuesta . . . . .	41
5.3. Análisis univariado . . . . .	43
5.4. Análisis multivariado . . . . .	45
5.5. Desarrollo, entrenamiento y validación de los modelos . . . . .	48
5.5.1. Ajuste de intercepto - regresión logística . . . . .	48
5.5.2. Análisis del umbral de decisión . . . . .	49
5.5.3. Validación de los modelos e iteración . . . . .	50
5.5.4. Validación <i>out time</i> . . . . .	51
<b>6. Resultados y análisis</b>	<b>53</b>
6.1. Regresión logística . . . . .	53
6.2. <i>Random forest</i> . . . . .	57
6.3. <i>Support vector machine</i> . . . . .	62
6.4. Tablas comparativas . . . . .	65
<b>7. Conclusiones y trabajo futuro</b>	<b>67</b>
<b>Bibliografía</b>	<b>69</b>

<b>Anexos</b>	<b>72</b>
<b>Anexo A. Análisis univariado</b>	<b>72</b>
<b>Anexo B. Construcción de características</b>	<b>76</b>