

Tabla de Contenido

1. Introducción	1
1.1. Contexto de Trabajo	1
1.1.1. Situación Actual y Oportunidad	1
1.1.2. Descripción y Justificación del Proyecto	2
1.2. Objetivos	4
1.2.1. Objetivo General	4
1.2.2. Objetivos Específicos	4
1.3. Hipótesis de Investigación	5
1.4. Metodología	5
1.5. Resultados Esperados	6
2. Marco Conceptual	7
2.1. World Wide Web	7
2.2. Web 2.0	8
2.3. Extracción de datos	8
2.3.1. Application Programming Interfaces (API's)	9
2.3.2. Web Crawlers	9
2.4. Data Mining y Web Mining	9
2.5. Text Mining	10
2.5.1. Unsupervised Learning	11
2.5.2. Supervised Learning	11
2.6. Topic Modelling	11
2.7. Sentiment Analysis	12
2.7.1. Planteamiento del Problema	13
2.7.2. Diferentes Niveles	13
2.7.3. Diferentes Enfoques	14
2.8. Teoría de Grafos	14
2.8.1. Representación Matricial	15
3. Consumo de Marihuana	16
3.1. Estudio de Predictores	16
3.1.1. Intrapersonal	16
3.1.2. Cultural o Referente a Actitudes	18
3.1.3. Social o interpersonal	19
3.1.4. Otras variables	21
3.2. Entorno Social	23

3.2.1.	Incrustación Social	23
3.2.2.	Estatus Social	24
3.2.3.	Proximidad Social a Consumidores	24
4.	Diseño	25
4.1.	Requerimientos	25
4.1.1.	Variables Originales	25
4.1.2.	Segmentación Original	26
4.1.3.	Selección de métricas y segmentación	27
4.1.4.	Indicadores Finales	27
4.2.	Descripción de Datos	28
4.2.1.	Datos disponibles	29
4.2.2.	Estructura Necesaria	31
4.2.3.	Etiquetado de <i>Tweets</i>	32
4.2.4.	Etiquetado de Usuarios	32
4.3.	Evaluación de Rendimiento	33
4.4.	Diseño de Aplicación	34
4.4.1.	Recolección de Datos	35
4.4.2.	Inteligencia de <i>Tweets</i>	40
4.4.3.	Inteligencia de Usuarios	41
4.4.4.	Visualización de Resultados	44
4.4.5.	Evaluación de Rendimiento	47
4.4.6.	Mantenimiento de Datos	48
5.	Implementación	53
5.1.	Herramientas de Desarrollo	53
5.2.	Selección de Palabras Clave	55
5.3.	Etiquetado	55
5.3.1.	Etiquetado de <i>Tweets</i>	56
5.3.2.	Etiquetado de Usuarios	56
5.4.	Implementación de la Aplicación	56
5.4.1.	Módulo de Recolección de Datos	56
5.4.2.	Módulo de Mantenimiento de Datos	58
5.4.3.	Módulo de Inteligencia de <i>Tweets</i>	59
5.4.4.	Módulo de Inteligencia de Usuarios	62
5.4.5.	Módulo de Evaluación de Rendimiento	65
5.4.6.	Módulo de Visualización de Resultados	66
6.	Resultados	69
6.1.	Palabras Clave	69
6.2.	Recolección de Datos de <i>Twitter</i>	70
6.3.	Etiquetado	73
6.3.1.	Etiquetado de <i>Tweets</i>	74
6.3.2.	Encuesta a Usuarios	75
6.4.	Evaluación de Algoritmos	76
6.4.1.	Consumo en <i>tweets</i>	76
6.4.2.	Políticas en <i>tweets</i>	77

6.4.3.	Edad de Usuarios	78
6.4.4.	Consumo de Usuarios	79
6.5.	Métricas	81
6.5.1.	Prevalencia	81
6.5.2.	Frecuencia de Consumo	82
6.5.3.	Polaridad	83
6.5.4.	Polaridad de Políticas	85
6.5.5.	Amigos Consumidores	87
6.5.6.	Oferta de Marihuana	87
6.5.7.	Palabras Utilizadas	87
7.	Conclusiones	89
7.1.	Conclusiones Generales	89
7.2.	Trabajo Futuro	90
	Bibliografía	91
	A. Proceso de validación de ingreso	99
	B. Vistas de prototipo funcional	102

Índice de tablas

4.1. Datos del usuario	30
4.2. Datos del <i>tweets</i>	30
4.3. Etiquetado de <i>tweets</i>	31
4.4. Etiquetado de usuarios	32
4.5. Matriz de Confusión	34
4.6. Restricciones de la API de Twitter	36
6.1. Palabras Clave La etiqueta (f) indica la aplicación de la regla	70
6.2. Medidas de Acuerdo	74
6.3. Heterogeneidad en las etiquetas	75
6.4. Rendimiento de <i>Naive Bayes</i> para el consumo en <i>tweets</i>	77
6.5. Rendimiento de <i>Voted Perceptron</i> para el consumo en <i>tweets</i>	77
6.6. Rendimiento de SVM para el consumo en <i>tweets</i>	77
6.7. Rendimiento de SVM para políticas en <i>tweets</i>	78
6.8. Rendimiento de <i>Voted Perceptron</i> para políticas en <i>tweets</i>	78
6.9. Rendimiento de C4.5 para políticas en <i>tweets</i>	78
6.10. Rendimiento de algoritmos de edad	79
6.11. Rendimiento de SVM para consumo en usuarios	80
6.12. Influencia de variables en el consumo de marihuana	80
6.13. Presencia de palabras en <i>tweets</i> de consumo	88
6.14. Presencia de palabras en <i>tweets</i> de políticas	88

Índice de figuras

1.1.	Regresión lineal de prevalencia Fuente: Elaboración Propia	3
1.2.	Sensibilidad de costos (prevalencia) Fuente: Elaboración Propia	4
4.1.	Diagrama Modular de la Aplicación Fuente: Elaboración Propia	35
4.2.	Casos de uso de MCD Fuente: Elaboración Propia	36
4.3.	Cola de Credenciales Fuente: Elaboración Propia	36
4.4.	Crawler de Usuarios Fuente: Elaboración Propia	38
4.5.	Ejemplo de Iteración Fuente: Elaboración Propia	38
4.6.	Casos de uso de MIT Fuente: Elaboración Propia	40
4.7.	Casos de uso de MIU Fuente: Elaboración Propia	42
4.8.	Casos de uso de MVR Fuente: Elaboración Propia	44
4.9.	Gráfico de polaridad Fuente: Elaboración Propia	45
4.10.	Gráfico de polaridad promedio Fuente: Elaboración Propia	46
4.11.	Gráfico de total de polaridad Fuente: Elaboración Propia	46
4.12.	Gráfico de prevalencia Fuente: Elaboración Propia	46
4.13.	Nube de palabras de encuesta Fuente: Sitio TagCrowd.com	46
4.14.	Estructura Navegacional Fuente: Elaboración propia	47
4.15.	Casos de Uso de MER Fuente: Elaboración propia	47
4.16.	Casos de Uso de MMD Fuente: Elaboración propia	49
4.17.	Diagrama E-R Fuente: Elaboración propia	49
4.18.	Entidad Usuarios Fuente: Elaboración propia	50
4.19.	Entidad <i>Tweets</i> Fuente: Elaboración propia	51
4.20.	Entidad Keywords Fuente: Elaboración propia	51
4.21.	Entidad Cliente Fuente: Elaboración propia	51
5.1.	Diagrama UML de Clases del MCD Fuente: Elaboración Propia	57
5.2.	Modelo E-R de mdb Fuente: Elaboración Propia	59
5.3.	Modelo E-R de Tagging Fuente: Elaboración Propia	60
5.4.	Modelo E-R de Trace Fuente: Elaboración Propia	60
5.5.	Modelo de Relations Fuente: Elaboración Propia	61
5.6.	Diagrama UML de Clases del MMD Fuente: Elaboración Propia	61
5.7.	Diagrama UML de Clases del MIT Fuente: Elaboración Propia	63
5.8.	Diagrama UML de Clases del MIU Fuente: Elaboración Propia	65
5.9.	Diagrama UML de Clases del MER Fuente: Elaboración Propia	67
6.1.	Gráfico de usuarios acumulados Fuente: Elaboración Propia	71
6.2.	Gráfico de cuentas acumuladas Fuente: Elaboración Propia	71

6.3. Número de <i>tweets</i> por año Fuente: Elaboración Propia	72
6.4. Curva de Lorenz de <i>Tweets</i> Fuente: Elaboración Propia	72
6.5. Distribución de Amigos Fuente: Elaboración Propia	73
6.6. Distribución de Seguidores Fuente: Elaboración Propia	74
6.7. Distribución de edad de la muestra Fuente: Elaboración Propia	75
6.8. Evaluación gráfica de SVM Fuente: Elaboración Propia	79
6.9. Prevalencia Anual Fuente: Elaboración Propia	82
6.10. Prevalencia Nacional Fuente: Estudio Nacional de Drogas	82
6.11. Comparación de Prevalencia Fuente: Elaboración Propia	83
6.12. <i>Tweets</i> de Consumo por Usuario Fuente: Elaboración Propia	83
6.13. Frecuencia de Consumo Fuente: Elaboración Propia	84
6.14. Polaridad de <i>Tweets</i> Fuente: Elaboración Propia	84
6.15. Polaridad de Usuarios Fuente: Elaboración Propia	85
6.16. Comparación de Percep. de Riesgo Fuente: Elaboración Propia	86
6.17. Polaridad en <i>Tweets</i> de Políticas Fuente: Elaboración Propia	86
6.18. Vecindario Consumidor Fuente: Elaboración Propia	87
A.1. Proceso BPMN de ingreso (1) Fuente: Elaboración Propia	100
A.2. Proceso BPMN de ingreso (2) Fuente: Elaboración Propia	101
B.1. Vista de login Fuente: Elaboración Propia	102
B.2. Vista de polaridad anual Fuente: Elaboración Propia	103
B.3. Vista de polaridad mensual Fuente: Elaboración Propia	104
B.4. Vista de polaridad diaria Fuente: Elaboración Propia	105
B.5. Vista de prevalencia Fuente: Elaboración Propia	106
B.6. Vista de polaridad de políticas Fuente: Elaboración Propia	107