

Tabla de Contenido

1. Introducción	1
1.1. Identificación y Formulación del Problema	2
1.2. Objetivos del Trabajo de Título	2
2. Marco Teórico y Estado del Arte	4
2.1. Machine Learning	4
2.1.1. Aprendizaje supervisado	4
2.1.2. Aprendizaje no supervisado	6
2.1.2.1. Latent Dirichlet Allocation (LDA)	7
2.1.2.2. Distributed Dictionary Representations (DDR)	8
2.2. Tópicos	10
2.3. Trabajos similares	12
3. Datos	14
3.1. Datos	14
4. Diseño, Visualización y Métricas	19
4.1. Tópicos	19
4.2. Técnicas utilizadas	20
4.3. Visualización	21
4.4. Métricas	22
5. Resultados	26
5.1. Distributed Dictionary Representations (DDR)	26
5.1.1. Validación del método utilizado	26
5.1.1.1. Ejemplos de clasificaciones realizadas	27
5.1.2. Distribución general de los tópicos	30
5.1.3. Comportamiento de las cuentas	30
5.1.4. Ranking de tópicos y RCA	32
5.1.5. Dendrograma, proximidad y agrupación de cuentas	33
6. Análisis de los Resultados	38
6.1. DDR	38
6.1.1. Encuesta y precisión del modelo	38
6.1.2. Mejoras al modelo utilizado	38
6.1.3. Clasificación de tópicos	39
6.1.3.1. Comportamiento de las cuentas	40
6.1.3.2. Cantidad de tweets por cuenta	41

6.1.3.3. Matriz de proximidad y agrupamiento de cuentas	41
7. Conclusión	45
7.1. Metodología	45
7.2. Validación del modelo	46
7.3. Cobertura de tópicos	46
7.4. Ranking y proximidad	46
7.5. Trabajo a futuro	47
Bibliografía	48
Anexos	53
A. Tópicos no utilizados/fusionados	53
B. Diccionarios utilizados	53
B.1. Arte, Cultura, Entretenimiento y Medios	54
B.2. Catástrofes y accidentes	54
B.3. Ciencia y tecnología	55
B.4. Conflicto, guerra y paz	55
B.5. Deporte	56
B.6. Economía, negocios y finanzas	56
B.7. Educación	57
B.8. Estilo de vida y tiempo libre	57
B.9. Interés humano, animales, insólito	57
B.10. Mano de obra	58
B.11. Medio ambiente	58
B.12. Meteorología	59
B.13. Policía y justicia	59
B.14. Política	60
B.15. Religión y culto	61
B.16. Salud	61
B.17. Sociedad	62
C. Latent Dirichlet Allocation (LDA)	62
D. Análisis de resultados - LDA	65
E. Ejemplos de gráficos de cuentas	66
F. Tablas de RCA de los tópicos utilizados	68