

Tabla de Contenidos

1. Introducción	1
1.1. Motivación y Antecedentes	1
1.2. Descripción del problema	3
1.3. Objetivos	4
1.3.1. Objetivos Generales	4
1.3.2. Objetivos Específicos	4
1.4. Aporte de la memoria	4
1.4.1. Alcances	4
1.5. Estructura de la Memoria	5
2. Marco Teórico	6
2.1. Áreas del conocimiento	7
2.1.1. Machine Learning	7
2.1.2. Procesamiento del Lenguaje Natural	8
2.2. Redes Neuronales Artificiales	9
2.3. Word Embeddings	11
2.3.1. Word2Vec	12
2.4. Limitaciones de los Word Embeddings	15
2.4.1. Definición de Similitud	15
2.4.2. Ovejas Negras	15
2.4.3. Antónimos	15
2.4.4. Sesgo del Corpus	16
2.4.5. Falta de Contexto	16
2.5. Aplicaciones con Word Embedding	16
2.5.1. Similitud de Word-Embeddings	17
2.5.2. Distancia Euclidiana	17
2.5.3. Similitud de Coseno	17
2.5.4. Distancia Euclidiana v/s Similitud de Coseno	18
2.5.5. Clasificación	18
2.5.5.1. Regresión Logística	18
2.5.5.2. Support Vector Machine	20
2.5.5.3. Random Forest	20
2.5.6. Métrica de Sesgo	20
2.5.7. Reducción de Dimensionalidad	20
2.5.7.1. UMAP: <i>Uniform Manifold Approximation and Projection</i> . .	21
3. Metodología	22

3.1.	Base de Datos	22
3.2.	Resumen de Metodología Propuesta	23
3.3.	Desarrollo de las Metodologías Propuesta	24
3.3.1.	Pre-procesamiento de Datos	24
3.3.1.1.	Exploración de los Datos	24
3.3.1.2.	Tokenización de los Documentos	24
3.3.1.3.	Normalización de las Palabras	24
3.3.1.4.	Desambiguación de las palabras	25
3.3.1.5.	Unión de los Corpus	26
3.3.2.	Obtención de Word Embeddings	26
3.3.3.	Comparación de los Word Embeddings	27
3.3.3.1.	Similitud de Semántica entre Países	27
3.3.3.2.	Evolución de la Semántica para cada País	28
3.3.3.3.	Matriz de Evolución Semántica	28
3.3.3.4.	Matriz de Evolución Semántica	28
3.3.4.	Relación de Grupos Sociales a Conceptos Claves de Emprendimiento	29
3.3.5.	Predicción del Sentimiento Asociado a Emprendimiento	30
3.3.6.	Medición de Sesgo a Través de RND	30
4.	Resultados y Análisis	31
4.1.	Interpretabilidad de Emprendimiento para Diferentes Países	31
4.2.	Evolución de la Semántica para Cada uno de los Países	32
4.2.1.	Evolución Semántica de Estados Unidos	32
4.2.2.	Evolución Semántica de Nueva Zelanda	34
4.2.3.	Evolución Semántica de Reino Unido	34
4.2.4.	Evolución Semántica de Irlanda	35
4.2.5.	Evolución Semántica de Canadá	36
4.2.6.	Evolución Semántica de Australia	37
4.3.	Evolución de Emprendimientos Utilizando Palabras Anclas	38
4.3.1.	Evolución de Emprendimiento	42
4.4.	Relación de grupos sociales a conceptos claves de emprendimiento	44
4.4.1.	Similitud Observada para Términos de Genero	44
4.4.2.	Similitud Observada para Términos de Ciudadanía	45
4.4.3.	Similitud Observada para Religiones	46
4.4.4.	Similitud Observada para Nombres Relacionados a Diferentes Etnias	48
4.5.	Medición de Sesgo a Través de RND	51
4.6.	Predicción del Sentimiento Asociado a Emprendimiento a lo Largo de los Años	54
4.7.	Resumen y Caracterización de los Países	56
4.7.1.	Caracterización de Estados Unidos	56
4.7.2.	Caracterización de Nueva Zelanda	56
4.7.3.	Caracterización de Reino Unido	56
4.7.4.	Caracterización de Irlanda	57
4.7.5.	Caracterización de Canadá	57
4.7.6.	Caracterización de Australia	57
5.	Conclusiones	58
	Bibliografía	60

