

Tabla de Contenido

1. Introducción	1
1.1. Objetivo General	2
1.2. Objetivos Específicos	3
2. Marco de trabajo	4
2.1. Utilidad de un agente	4
2.2. Tipos de agentes	5
2.3. Resultados actuales	6
2.4. Conceptos	7
2.4.1. Procesamiento de texto	7
2.4.2. Machine learning	8
2.5. Tecnologías disponibles	10
3. Solución propuesta	12
3.1. Sistema actual	12
3.2. Etapa inicial	12
3.3. Etapa final	13
4. Tratamiento de los datos	14
4.1. Pasos realizados	15
5. Base de Datos	17
5.1. Motor de base de datos	17

5.2. Modelo de datos	17
6. Desarrollo de la API	23
6.1. Versiones de metodologías	24
6.2. Mantenedores	24
6.2.1. Mantenedor de frases	24
6.2.2. Mantenedor de etiquetas	25
6.2.3. Mantenedor de tipos de mensaje	25
6.2.4. Mantenedor de Modelos	25
6.3. Seguridad	26
7. Métodos utilizados	27
7.1. Expresiones regulares	27
7.2. TF-IDF	27
7.3. Multiclass classification	27
8. Interfaces del sistema	29
9. Evaluación de los resultados	34
9.1. Preprocesamiento de datos	34
9.1.1. Separación de los datos	34
9.2. Entrenamiento y evaluación	35
9.2.1. Primer modelo: <i>One-hot encoding</i> usando <i>stemming</i>	35
9.2.2. Segundo modelo: <i>One-hot encoding</i> sin uso de <i>stemming</i>	37
9.2.3. Tercer modelo: Reducción de clases, <i>One-hot encoding</i> usando <i>stemming</i>	38
9.2.4. Cuarto modelo: TF-IDF usando <i>stemming</i>	40
9.2.5. Quinto modelo: Reducción de clases, TF-IDF usando <i>stemming</i>	42
9.2.6. Sexto modelo: Ejemplo de sobreajuste por sobremuestreo	43
9.2.7. Resumen de modelos	45
9.2.8. Validación cruzada	45

10. Conclusión	48
10.1. Mejoras a futuro	49
Bibliografía	52