

Tabla de Contenido

1. Introducción	1
2. Antecedentes	3
2.1. Marco teórico	3
2.1.1. Aprendizaje de máquina	3
Hiperparámetros	4
Step	4
Batch size	4
Epoca	4
Acumulación de gradiente	4
Learning rate	4
Warm up	4
Softmax	5
Representación de palabras para modelos de PLN	5
Conjunto de Datos	5
2.1.2. Métricas de evaluación	5
2.2. Estado del arte	6
2.2.1. AutoAtención	6
2.2.2. Pre entrenamiento y entrenamiento específico.	8
2.2.3. BERT: Bidirectional Encoder Representations from Transformers	9
2.2.4. ALBERT: A Lite BERT	10
2.2.5. DistilBERT	11
2.2.6. GLUES	12
Natural Language Inference: XNLI	12
Paraphrasing: PAWS-X	12
Named Entity Recognition: NER	12
Part-of-Speech Tagging: POS	13
Document Classification: MLDoc	13
Question Answering: MLQA, TAR and XQuAD	13
3. Problema	14
3.1. Objetivos	16
3.1.1. Objetivo General	16
3.1.2. Objetivos Específicos	16
4. Solución	17
4.1. Evaluación de modelos s-ALBERT	17
4.2. Pre entrenamiento y evaluación de DistilBETO	18

4.3. Evaluación en Hardware de común acceso	19
5. Resultados y análisis	20
5.1. Evaluación en GLUES	20
Tiempo de entrenamiento específico	20
Evaluación en GLUES	21
DistilBETO	23
5.2. Rendimiento en Hardware	24
6. Conclusiones	25
Bibliografía	27
Anexo A.	30
A.1. Hiperparámetros	30
A.2. Tiempos de inferencia	33