**Modelos BERT para Clasificación de Emociones en Español**

BERT (Bidirectional Encoder Representations from Transformers) es un modelo de lenguaje basado en la arquitectura Transformer, preentrenado con tareas como Masked Language Modeling y Next Sentence Prediction. BERT permite representar contextos bidireccionales en el texto, lo cual lo hace eficaz para tareas de clasificación de emociones.

**Modelos BERT relevantes para el idioma español**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Modelo** | **Tamaño** | **Entrenado en** | **Casing** | **Tokenizer** | **Arquitectura** | **Observaciones** |
| bert-base-multilingual-cased | 110M | 104 idiomas | Sí | WordPiece | BERT-base | Multilingüe, buena cobertura, menor precisión que modelos específicos. |
| dccuchile/bert-base-spanish-wwm-cased | 110M | Wikipedia, OPUS, Oscar (español) | Sí | WordPiece | BERT-base + WWM | Modelo monolingüe, con enmascaramiento de palabra completa. |
| bertin-project/bertin-roberta-base-spanish | ~125M | Oscar corpus | No | Byte-level BPE | RoBERTa-base | Basado en RoBERTa, mejor rendimiento; no tiene NSP. |
| PlanTL-GOB-ES/roberta-base-bne | ~125M | Boletín Oficial del Estado | No | Byte-level BPE | RoBERTa-base | Enfocado en lenguaje formal/jurídico. |
| pysentimiento/bert-base-spanish-uncased | 110M | Español general | No | WordPiece | BERT-base | Optimizado para análisis de emociones y sentimientos. |
| BETO | 110M | Wikipedia en español | Sí | WordPiece | BERT-base | Versión clásica, disponible en cased y uncased. |
| finiteautomata/beto-sentiment-analysis | 110M | BETO fine-tuned | Sí | WordPiece | BERT-base | Fine-tuned para emociones y sentimientos. |

**Comparativa entre modelos destacados**

|  |  |  |  |
| --- | --- | --- | --- |
| **Criterio** | **bert-base-multilingual-cased** | **dccuchile/bert-base-spanish-wwm-cased** | **pysentimiento/bert-base-spanish-uncased** |
| Cobertura | Amplia (multilingüe) | Español nativo | Español nativo |
| Precisión | Moderada | Alta | Alta |
| Casing | Sí | Sí | No |
| WWM | No | Sí | No |
| Fine-tuned en emociones | No | No | Sí |
| Tokenizer | WordPiece | WordPiece | WordPiece |

**Recomendaciones para textos en español mexicano**

Dado que los textos provienen de estudiantes universitarios en México, se recomienda:

1. pysentimiento/bert-base-spanish-uncased: Fine-tuned para emociones, alto rendimiento.  
2. dccuchile/bert-base-spanish-wwm-cased: Español latinoamericano, requiere fine-tuning.  
3. bert-base-multilingual-cased: Uso general, menos específico.  
4. finiteautomata/beto-sentiment-analysis: Ya adaptado a tareas afectivas.