

# Sequence Classification

Presented by:  
Majd, Nada, Nujud, Abdulmlak  
Almuhnna

# Main Goal

The objective of this project is to develop a model capable of classifying sequences into predefined categories based on learned patterns and features from the data.

# Methodology

Our methodology involves fine-tuning one model using cleaned data and another model without any data cleaning.

# Model Initialization:

BERT-based Architecture:

The model is initialized using the pre-trained checkpoint `aubmindlab/bert-base-arabertv02`.

We specified the number of output labels based on the dataset (3 labels), ensuring the model fits the classification task.

# Without fine tuning

```
from transformers import pipeline
```

```
symantic_analyser = pipeline("sentiment-analysis", model="aubmindlab/bert-base-arabertv02")
```

```
text = 'قدم خدمات الرعاية الصحية الأولية إلى مراجعيها عبر 18 مركزاً صحياً تنتشر جغرافياً، في مختلف أنحاء الإمارة'
```

```
result = symantic_analyser(text)
```

```
result
```

```
Some weights of BertForSequenceClassification were not initialized from the model checkpoint at aubmindlab/bert-base-arabertv02. You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.  
[{'label': 'LABEL_1', 'score': 0.7252967357635498}]
```

```
text2 = 'المال'
```

```
r = symantic_analyser(text2)
```

```
r
```

```
[{'label': 'LABEL_1', 'score': 0.6437215209007263}]
```

The image features a white background with four decorative purple circles in the corners. The top-left circle is a solid dark purple. The top-right circle is a solid dark purple. The bottom-left corner contains two overlapping circles: a larger dark purple one and a smaller, lighter purple one. The text "With fine tuning" is centered in the middle of the image.

**With fine tuning**

# Training Arguments:

## Training Setup:

- We defined the model's training parameters including:
  - Learning rate:  $4e-5$
  - Batch size: 4
  - Number of epochs: 1

# Trainer Initialization and Training:

## Model Training:

- We utilized the Trainer class from transformers to handle training.
- After running for 3000 steps, the model achieved a training loss of 0.0206, showing it successfully learned from the data.



# Performing Inference

```
from transformers import pipeline
```

```
model_inf = pipeline("text-classification", model = "/content/mymodel",
tokenizer="/content/mytokenizer", return_all_scores=False)
```

```
example = ""
```

،رأس الخيمة: «الخليج» افتتح الدكتور عبدالله النعيمي، مدير منطقة رأس الخيمة الطبية،مركز الظيث للرعاية الصحية الأولية،يخدم منطقة حيوية تضم كثافة سكانية عالية. وأوضح د. النعيمي، خلال زيارته للمركز الصحي، أن أعمال الصيانة للمركز بدأت في يناير/كانون الثاني من العام الماضي، واشتملت على تعديلات في المبنى وتطوير مرافقه وخدماته،من أهمها إنشاء حجرات إضافية، وأقسام جديدة للمختبر والتمريض والعيادات وتجهيز المواقع الخارجية لسيارات المراجعين والموظفين، وتعديلات داخلية في المبنى الرئيسي للمركز الصحي الحيوي.

وبين النعيمي أن الهدف من مشروع توسعة مركز الظيث وتطويره هو الارتقاء،بالخدمات الصحية المقدمة لأهالي المنطقة، وتسهيل تنقل المراجعين بين مرافق المركز وتسهيل عملية دخول المراجعين بسياراتهم إلى حرم المركز. وأضاف النعيمي «الأشغال العامة» تسلمت المبنى من المنطقة الطبية في وقت سابق، وتكفلت بتنفيذ التوسعات. والتعديلات المنشودة، ليكون المركز جاهزاً للعمل في إبريل/نيسان الحالي وأشار الدكتور النعيمي إلى أن المنطقة تقدم خدمات الرعاية الصحية الأولية. إلى مراجعيها عبر 18 مركزاً صحياً تنتشر جغرافياً، في مختلف أنحاء الإمارة.

```
"""
```

```
result = model_inf(example)
print(result)
```

```
[{'label': 'LABEL_2', 'score': 0.9999425411224365}]
```

```
example = ""
```

الحمادي وعبدالله الظاهري وخالد علي ربيع ومحمد الحمادي ومنال بن عمرو ومروان عبدالله صالح و هم جميعاً من العاملين في مجال الإخراج وقد تتلمذوا على أيدي كوكبة من المخرجين الإماراتيين من الجيل المفصل وقد برزوا من خلال أعمالهم المسرحية التي تدور حول الأرض والوطن والهوية بدلالات ورموز مختلفة في أبعادها الفنية والفكرية، والموقف من مواكبة العصر أو التمسك بالماضي وأحلام الإنسان العادي وطموحاته في هذه الحياة التي تعج بالصراعات الاجتماعية والسياسية والاقتصادية.

ولا تقتصر فعاليات المهرجان على العروض وحسب، وإنما أدرجت اللجنة الاستشارية للمهرج ان وفي كل دورة، تكريم أحد الفنانين المسرحيين من الجيل

المؤسس في الدولة، حيث يكرم في هذه الدورة الفنان الشامل عبدالله صالح، وهي لفئة نبيلة

تحتسب لإدارة المهرجان.نواف يونس

```
"""
```

```
result = model_inf(example)
print(result)
```

```
[{'label': 'LABEL_0', 'score': 0.9999538660049438}]
```

```
example = '''
```

على مستوى 6. 30 درهم مسجلاً خسارة بنسبة 10% من خلال تداول 30 ألف سهم بقيمة 189 ألف درهم تلاه سهم "الشركة الوطنية للتأمينات العامة" الذي انخفض بنسب .ة 37. 8% ليغلق على مستوى 3. 94 درهم من خلال تداول 300 سهم بقيمة 1182 درهماً ومنذ بداية العام بلغت نسبة الارتفاع في مؤشر سوق الإمارات المالي. وبلغ إجمالي قيمة التداول 245. 48 مليار درهم ،25%. 70 وبلغ عدد الشركات التي حققت ارتفاعاً سعرياً 70 من أصل . وعدد الشركات المتراجعة 35 شركة 120،

```
'''
```

```
result = model_inf(example)
print(result)
```

```
[{'label': 'LABEL_1', 'score': 0.9999779462814331}]|
```

The image features a white background with four decorative purple circles in the corners. The top-left circle is a solid purple. The top-right circle is a solid purple. The bottom-left corner contains two overlapping circles: a larger, darker purple one and a smaller, lighter purple one. The text "With Cleaning" is centered in the middle of the image.

**With Cleaning**

# Performing Inference(after cleaning):

```
from transformers import pipeline
```

```
model_inf = pipeline("text-classification", model = "/content/mymodel",  
tokenizer="/content/mytokenizer", return_all_scores=False)
```

```
example = ""
```

```
رأس الخيمة: «الخليج» افتتح الدكتور عبدالله النعيمي، مدير منطقة رأس الخيمة الطبية  
،مركز الطيت للرعاية الصحية الأولية  
مي، خلال زيارته للمركز الصحي، أن أعمال الصيانة للمركز بدأت في يناير/كانون الثاني  
اشتملت على تعديلات في المبنى وتطوير مرافقه وخدماته، من أهمها إنشاء حجرات إضافية  
لمختبر والتمريض والعيادات، وتجهيز المواقع الخارجية لسيارات المراجعين والموظفين  
،وتعديلات داخلية في المبنى الرئيسي للمركز الصحي الحيوي  
وبين النعيمي أن الهدف من مشروع توسعة مركز الطيت وتطويره هو الارتقاء  
،بالخدمات الصحية المقدمة لأهالي المنطقة، وتسهيل تنقل المراجعين بين مرافق المركز  
،وتسهيل عملية دخول المراجعين بسياراتهم إلى حرم المركز  
وأضاف النعيمي أن «الأشغال العامة» تسلمت المبنى من ا  
لمنطقة الطبية في وقت سابق، وتكفلت بتنفيذ التوسعات والتعديلات المنشودة  
ليكون المركز جاهزاً للعمل في إبريل/نيسان الحالي. وأشار الدكتور النعيمي إلى  
أن المنطقة تقدم خدمات الرعاية الصحية الأولية إلى مراجعيها
```

```
عبر 18 مركزاً صحياً تنتشر جغرافياً، في مختلف أنحاء الإمارة
```

```
"""
```

```
clean_example=cleaning(example)  
result = model_inf(clean_example)  
print(result)
```

```
[{'label': 'LABEL_2', 'score': 0.9998788833618164}]
```

```
example = '''
```

```
1. 17 درهم مرتفعاً بنسبة 9. 35% من خلال تداول 12. 05 مليون سهم بقيمة 14. 21 مليون درهم  
وجاء في المركز الثاني من حيث الارتفاع السعري سهم  
العربية للطيران" ليغلق على مستوى 1. 58 درهم مرتفعاً بنسبة"  
من خلال تداول 288 مليون سهم بقيمة 449. 32 مليون درهم 48. 7%  
"وسجل سهم "شركة أبوظبي الوطنية للتأمين.  
أكثر انخفاض سعري في جلسة التداول حيث أقفل سعر السهم  
على مستوى 6. 30 درهم مسجلاً خسارة بنسبة  
من خلال تداول 30 ألف سهم بقيمة 189 ألف درهم . تلاه سهم 10%  
الشركة الوطنية للتأمينات العامة" الذي انخفض بنسبة"  
ليغلق على مستوى 3. 94 درهم من خلال تداول 300 سهم بقيمة 1182 درهماً 37. 8%  
،%ومنذ بداية العام بلغت نسبة الارتفاع في مؤشر سوق الإمارات المالي 25. 70.  
وبلغ إجمالي قيمة التداول 245. 48 مليار درهم . وبلغ عدد الشركات التي  
حققت ارتفاعاً سعرياً 70 من أصل 120 وعدد الشركات المتراجعة 35 شركة
```

```
'''
```

```
clean_example=cleaning(example)  
result = model_inf(clean_example)  
print(result)
```

```
[{'label': 'LABEL_1', 'score': 0.9999696016311646}]
```

# Performing Accuracy on Inference:

```
from sklearn.metrics import accuracy_score
true=[]
pred=[]
for row in test_dataset:
    true.append(row.get('label'))

for row in test_dataset:
    pred.append(model_inf(test_dataset.get('text')))
acc=accuracy_score(true,pred)
acc
```

97.35

] Start coding or generate with AI.

The background features four large, solid purple circles positioned at the corners of the slide: top-left, top-right, bottom-left, and bottom-right. The text "Thank You" is centered in the middle of the slide.

**Thank You**