model\_type: llm

base\_model: {model\_path}

input\_features:

  - name: DOMANDE

    type: text

    preprocessing:

      truncation: true

      max\_sequence\_length: 210

      lowercase: true

output\_features:

  - name: RISPOSTE

    type: text

    preprocessing:

      max\_sequence\_length: 1000

      max\_length: 1000

      truncation: true

    metrics:

      - type: bert\_score

        model: dbmdz/bert-base-italian-uncased

      - type: loss

      ### - type: rouge, - type: perplexity

prompt:

  template: >-

    Sei un esperto cardiologo. Rispondi alla seguente domanda in modo empatico e con frasi complete e concise, terminando ogni frase con un punto.

    ### Esempio 1:

    Domanda: Quali sono i sintomi dell'ipertensione?

    Risposta: I sintomi dell'ipertensione includono mal di testa, vertigini e difficoltà respiratorie.

    ### Esempio 2:

    Domanda: Come posso controllare la mia pressione arteriosa?

    Risposta: Puoi controllare la tua pressione arteriosa con una dieta sana, esercizio fisico regolare e farmaci prescritti.

    ### Domanda: {DOMANDE}

    ### Risposta (Assicurati che la tua risposta termini con un punto e abbia senso compiuto):

generation:

  temperature: 0.1    ### per la generazione del testo deve essere molto basso

  max\_new\_tokens: 80

  top\_p: 0.9          ### aiuta a generare testo più coerente e fluido. , il modello avrà un equilibrio tra casualità e coerenza, garantendo che i token generati siano tra quelli con le probabilità più alte.

adapter:

  type: lora

quantization:

  bits: 4

preprocessing:

  split:

    type: random

    probabilities: [0.8, 0.1, 0.1]  ### a gestione di più sottoinsiemi può richiedere risorse aggiuntive.

  global\_max\_sequence\_length: 1210

trainer:

  type: finetune

  epochs: 4

  batch\_size: 2

  eval\_batch\_size: 2

  gradient\_accumulation\_steps: 8  # effective batch size = batch size \* gradient\_accumulation\_steps

  learning\_rate: 1.0e-4  # Prova con un learning rate leggermente più basso

  enable\_gradient\_checkpointing: true

  learning\_rate\_scheduler:

    decay: cosine

    warmup\_fraction: 0.1  # Aumenta il warmup fraction

    reduce\_on\_plateau: 0

### un checkpoint del modello verrà salvato ogni 500 step di addestramento, permettendo il ripristino del modello in caso di interruzioni.

early\_stopping:                             ### monitora la perdita di validazione (val\_loss) e interrompe l'addestramento se la perdita non migliora per 5 epoche consecutive, per prevenire l'overfitting (min: il monitoraggio è configurato per minimizzare la perdita).

  monitor: val\_loss

  patience: 5

  mode: min

logging:                                    ### i log dell'allenamento verranno inviati a TensorBoard.

  log\_to\_tensorboard: true

callbacks:

  - type: progress\_bar

  - type: model\_checkpoint

    save\_best\_only: true

    monitor: val\_loss

    mode: min

  - type: early\_stopping

    monitor: val\_loss

    patience: 5

    mode: min

Immagine che contiene linea, diagramma, Diagramma, testo

Descrizione generata automaticamente

Grafico 1: Perdita per Epoca

Asse X: Epoca

Asse Y: Perdita

Legenda:

Blu: Perdita di Addestramento

Arancione: Perdita di Validazione

Verde: Perdita di Test

**Analisi:**

**Perdita di Addestramento:**

La perdita di addestramento diminuisce costantemente dall'epoca 1 all'epoca 4.

La perdita iniziale di addestramento è circa 1.4 e si riduce a circa 0.8 alla quarta epoca.

**Perdita di Validazione:**

La perdita di validazione segue un trend simile alla perdita di addestramento, diminuendo da circa 1.3 a 0.9.

C'è un piccolo gap tra le perdite di addestramento e di validazione, indicando un buon adattamento.

**Perdita di Test:**

La perdita di test diminuisce anch'essa, ma mostra maggiore variabilità rispetto alle perdite di addestramento e validazione.

***La perdita iniziale di test è circa 1.2 e si riduce a circa 0.9.***

**Considerazioni:**

Il modello mostra un miglioramento consistente su tutti i set di dati di addestramento, validazione e test.

La diminuzione dei valori di perdita suggerisce che il modello sta apprendendo efficacemente nel corso delle epoche.

L'allineamento stretto tra le perdite di addestramento e di validazione indica un sovradattamento minimo, ovvero il modello sta mantenendo un buon equilibrio tra adattamento ai dati di addestramento e capacità di generalizzazione ai dati di validazione.

**Grafico 2: Perplessità per Epoca**

Asse X: Epoca

Asse Y: Perplessità

Legenda:

Blu: Perplessità di Addestramento

Arancione: Perplessità di Validazione

Verde: Perplessità di Test

**Analisi:**

**Perplessità di Addestramento:**

La perplessità di addestramento diminuisce leggermente dall'epoca 1 all'epoca 4, ma i valori rimangono piuttosto alti, partendo da circa 31993 a circa 31986.

**Perplessità di Validazione:**

La perplessità di validazione mostra un trend simile alla perplessità di addestramento, con valori che partono da circa 31992 a circa 31988.

**Perplessità di Test:**

La perplessità di test segue lo stesso trend, diminuendo leggermente ma mantenendo valori complessivamente alti.

**Considerazioni:**

Nonostante la riduzione, i valori alti di perplessità suggeriscono che il modello affronta ancora delle sfide nella previsione delle sequenze.

La coerenza tra le perplessità di addestramento, validazione e test indica una performance stabile del modello, anche se sono necessari miglioramenti.

**Grafico 3: Accuratezza dei Token per Epoca**

Asse X: Epoca

Asse Y: Accuratezza dei Token

Legenda:

Blu: Accuratezza dei Token di Addestramento

Arancione: Accuratezza dei Token di Validazione

Verde: Accuratezza dei Token di Test

**Analisi:**

**Accuratezza dei Token di Addestramento:**

L'accuratezza dei token di addestramento mostra un aumento dall'epoca 1 all'epoca 4.

Partendo da un valore molto basso, raggiunge circa 0.00012.

**Accuratezza dei Token di Validazione:**

L'accuratezza dei token di validazione aumenta anch'essa nel corso delle epoche, con valori vicini all'accuratezza dei token di addestramento.

**Accuratezza dei Token di Test:**

L'accuratezza dei token di test migliora allo stesso modo, raggiungendo un valore finale simile a quello delle accuratezze di addestramento e validazione.

**Considerazioni:**

L'aumento dell'accuratezza dei token su tutti i set di dati indica la capacità del modello di migliorare la precisione nella previsione dei singoli token.

L'allineamento stretto tra le accuratezze dei set di addestramento, validazione e test suggerisce una buona generalizzazione.

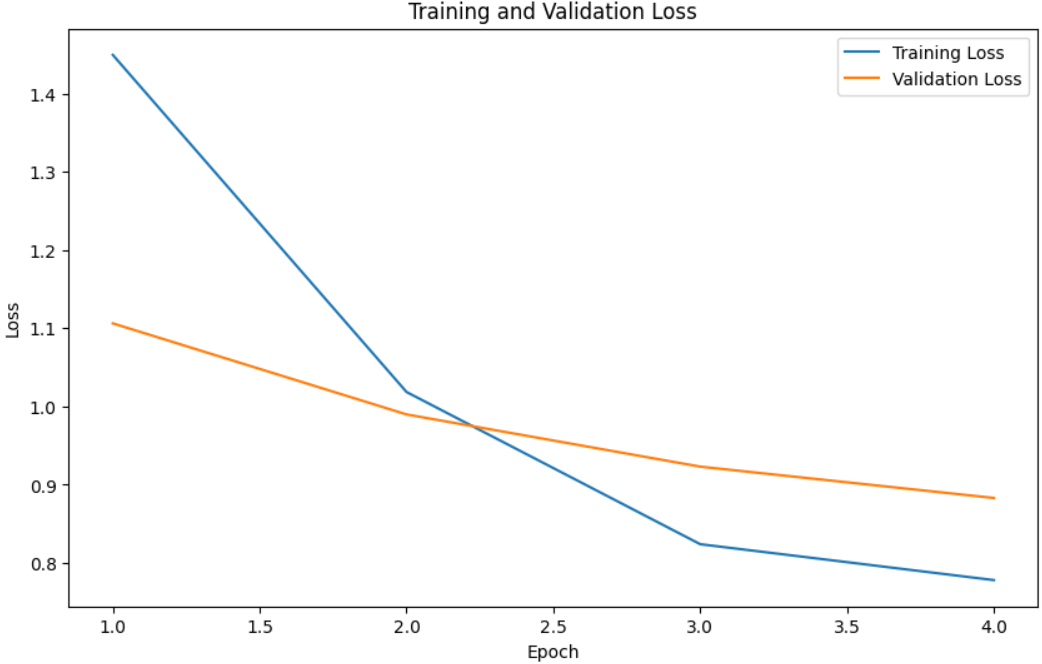
**Riepilogo Generale**

Performance del Modello: Il modello mostra miglioramenti su tutte le metriche nel corso delle epoche, indicando un apprendimento efficace.

Perdita: La riduzione costante dei valori di perdita su tutti i set di dati suggerisce che il modello sta minimizzando l'errore in modo efficace.

Perplessità: I valori alti di perplessità indicano potenziali problemi con la previsione delle sequenze, che richiedono ulteriori indagini.

Accuratezza dei Token: L'aumento dell'accuratezza dei token su tutti i set di dati dimostra un miglioramento nella precisione del modello.



**Grafico: Perdita di Addestramento e Validazione**

Asse X: Epoca

Asse Y: Perdita

Legenda:

Blu: Perdita di Addestramento

Arancione: Perdita di Validazione

**Analisi:**

**Perdita di Addestramento:**

La perdita di addestramento diminuisce costantemente dall'epoca 1 all'epoca 4.

Parte da un valore di circa 1.4 e si riduce a circa 0.8 alla quarta epoca.

La discesa è rapida nelle prime epoche e diventa più graduale successivamente.

**Perdita di Validazione:**

La perdita di validazione segue un trend simile alla perdita di addestramento, diminuendo da circa 1.1 a circa 0.9.

Anche in questo caso, la diminuzione è più rapida nelle prime epoche e poi diventa più graduale.

La perdita di validazione è costantemente inferiore alla perdita di addestramento fino alla terza epoca, dove diventano quasi parallele.

**Considerazioni:**

Miglioramento del Modello: Il modello mostra un miglioramento significativo con la riduzione costante della perdita sia per l'addestramento che per la validazione.

Buon Adattamento: La perdita di validazione inferiore a quella di addestramento nelle prime epoche suggerisce che il modello si sta adattando bene ai dati senza sovradattarsi eccessivamente.

Stabilità: La stabilità nella riduzione della perdita di validazione indica che il modello generalizza bene anche su dati non visti durante l'addestramento.

**Riepilogo**

Performance del Modello: Il modello sta migliorando efficacemente riducendo la perdita su entrambi i set di addestramento e validazione.

Validazione: La perdita di validazione inferiore e parallela a quella di addestramento suggerisce un buon equilibrio tra apprendimento e generalizzazione.

Questa analisi fornisce una visione completa del comportamento del modello in termini di perdita durante l'addestramento e la validazione.

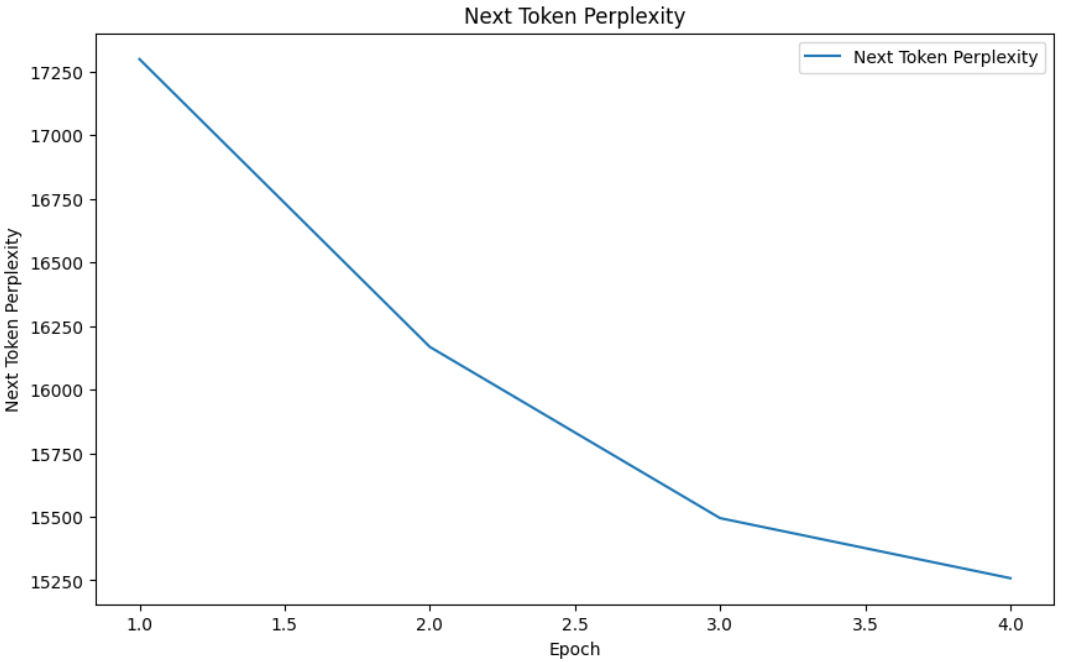


Grafico: Perplessità del Prossimo Token

Asse X: Epoca

Asse Y: Perplessità del Prossimo Token

Legenda: Next Token Perplexity

**Analisi:**

**Trend della Perplessità:**

La perplessità del prossimo token diminuisce costantemente dall'epoca 1 all'epoca 4.

Parte da un valore di circa 17250 e si riduce a circa 15250 alla quarta epoca.

La discesa è relativamente lineare, indicando un miglioramento costante ad ogni epoca.

**Considerazioni:**

Miglioramento del Modello: La diminuzione della perplessità indica che il modello sta migliorando la sua capacità di prevedere il prossimo token in una sequenza.

Efficienza dell'Addestramento: La riduzione costante e regolare suggerisce che l'addestramento sta procedendo efficacemente senza grossi problemi di sovradattamento o di sottodimensionamento.

Previsione delle Sequenze: La perplessità è una misura di quanto bene un modello prevede una sequenza. Valori più bassi indicano previsioni più accurate. La significativa riduzione della perplessità suggerisce che il modello sta apprendendo i pattern sottostanti nei dati.

**Riepilogo**

Performance del Modello: Il modello mostra un miglioramento significativo nella previsione del prossimo token, come indicato dalla diminuzione della perplessità.

Stabilità: La riduzione costante e regolare della perplessità attraverso le epoche suggerisce un processo di addestramento stabile ed efficace.

Potenziale: Continuando l'addestramento, il modello potrebbe ulteriormente migliorare la sua capacità predittiva, riducendo ulteriormente la perplessità.

Immagine che contiene testo, linea, diagramma, Diagramma

Descrizione generata automaticamente

Grafico: Accuratezza dei Token

Asse X: Epoca

Asse Y: Accuratezza dei Token

Legenda: Token Accuracy

**Analisi:**

**Trend dell'Accuratezza:**

L'accuratezza dei token mostra variazioni durante le epoche.

Parte da un valore di circa 4e-5 (0.00004) alla prima epoca.

Aumenta rapidamente fino a circa 7.5e-5 (0.000075) alla seconda epoca.

Diminuisce nuovamente a circa 6e-5 (0.00006) alla terza epoca.

Infine, risale a circa 7.5e-5 (0.000075) alla quarta epoca.

**Considerazioni:**

Variazioni: L'accuratezza dei token non segue un trend uniforme. C'è un aumento significativo nelle prime due epoche, seguito da una diminuzione e poi un nuovo aumento.

Efficienza dell'Addestramento: L'aumento dell'accuratezza fino alla seconda epoca indica un miglioramento nella capacità del modello di prevedere correttamente i token. La successiva diminuzione alla terza epoca potrebbe indicare un aggiustamento del modello che potrebbe aver avuto un impatto temporaneo sulle prestazioni.

Risalita: La ripresa dell'accuratezza alla quarta epoca suggerisce che il modello sta continuando ad apprendere e migliorare.

**Riepilogo**

Performance del Modello: Il modello mostra miglioramenti significativi nelle prime due epoche, una diminuzione temporanea nella terza, e un nuovo miglioramento nella quarta.

Variazioni: Le variazioni nell'accuratezza potrebbero essere dovute a diverse fasi di apprendimento del modello e possibili aggiustamenti dei parametri durante l'addestramento.