

## *Izveštaj o fine-tuningu DistilBERT modela za prepoznavanje imenovanih entiteta*

### 1. Uvod

U ovom eksperimentu, sproveo sam fine-tuning DistilBERT modela za zadatak prepoznavanja imenovanih entiteta (Named Entity Recognition - NER). Cilj je bio optimizovati performanse modela kroz eksperimentisanje sa različitim hiperparametrima. NER je ključni zadatak u obradi prirodnog jezika, koji ima široku primenu u ekstrahovanju informacija, analizi sentimenta i mnogim drugim oblastima.

### 2. Metodologija

2.1 Podaci Za ovaj eksperiment, odabrao sam CoNLL 2003 dataset, koji je široko priznat benchmark za NER zadatke. Ovaj dataset sadrži novinske članke sa oznakama za osobe, lokacije, organizacije i razne druge entitete, što ga čini idealnim za našu svrhu.

2.2 Priprema podataka Priprema podataka je bila ključna faza eksperimenta:

- Učitavanje datasea: Koristio sam Hugging Face datasets biblioteku za efikasno učitavanje CoNLL 2003 datasea.
- Tokenizacija: Implementirao sam posebnu funkciju za tokenizaciju koja ne samo da deli tekst na tokene, već i usklađuje oznake entiteta sa novonastalim tokenima. Ovo je kritičan korak jer BERT-baziran model kao što je DistilBERT koristi WordPiece tokenizaciju koja može podeliti reči na podreči.
- Primena tokenizacije: Primenio sam ovu funkciju na ceo dataset, pripremajući ga za trening.

2.3 Model Odabrao sam DistilBertForTokenClassification model, koji je već predtreiniran na velikom korpusu teksta. Ovaj model je inicijalizovan sa "distilbert-base-uncased" težinama, što nam omogućava da iskoristimo znanje stečeno tokom predtreninga i primenimo ga na naš specifični NER zadatak.

2.4 Trening Za proces treninga, oslonio sam se na Hugging Face Trainer API, koji pruža robustan okvir za fine-tuning transformerskih modela. Ključni elementi mog pristupa treningu uključuju:

- Data collator: Koristio sam DataCollatorForTokenClassification za efikasno procesiranje batch-eva, što je posebno važno za zadatke klasifikacije tokena.
- Metrike: Implementirao sam funkciju za računanje važnih metrika uključujući tačnost (accuracy), F1 score, preciznost i odziv. Ove metrike pružaju sveobuhvatan uvid u performanse modela.
- Early Stopping: Uključio sam EarlyStoppingCallback sa strpljenjem od 3 epohe kako bih sprečio prekomerno fitovanje modela.

## 2.5 Hiperparametri Eksperimentisao sam sa sledećim hiperparametrima:

- Learning rates: [1e-5, 3e-5, 5e-5] - Ovi opsezi su tipični za fine-tuning BERT-baziranih modela.
- Batch sizes: [16, 32] - Manji batch size-ovi su često efikasniji za fine-tuning, ali sam hteo da istražim i veće vrednosti.
- Broj epoha: [3, 5] - Ovo mi je omogućilo da istražim trade-off između dužine treninga i performansi.

## 3. Rezultati

Nakon temeljnog testiranja različitih kombinacija hiperparametara, najbolje performanse su postignute sa sledećom konfiguracijom:

- Learning rate: 5e-5
- Batch size: 16
- Broj epoha: 3

Ovaj model je postigao impresivan F1 score od 0.9883537364482007 na validacionom setu, što ukazuje na izuzetno visoku preciznost u prepoznavanju imenovanih entiteta.

## 4. Diskusija

Rezultati eksperimenta su izuzetno ohrabrujući. DistilBERT model je pokazao sposobnost da postigne state-of-the-art performanse na NER zadatku, i to sa relativno kratkim treningom. Posebno je zanimljivo primetiti da je model sa manjim batch size-om (16) i manjim brojem epoha (3) nadmašio modele sa većim batch size-om i dužim treningom.

Ovo otkriće sugerise nekoliko stvari:

1. Efikasnost transfera znanja: DistilBERT, iako "destilovan" i manji od originalnog BERT-a, zadržava dovoljno predtreirane informacije da brzo konvergira na specifičnom zadatku.
2. Važnost learning rate-a: Viši learning rate od 5e-5 se pokazao najefikasnijim, možda omogućavajući modelu da brže pronađe optimum za ovaj zadatak.
3. Prednost manjih batch-eva: Manji batch size od 16 je dao bolje rezultate, što može ukazivati na bolju generalizaciju zbog veće stohastičnosti u procesu učenja.
4. Efikasnost kratkog treninga: Činjenica da je 3 epohe bilo dovoljno za postizanje vrhunskih rezultata ukazuje na brzu adaptabilnost modela i efikasnost transfer learning pristupa.
5. Zaključak

Uspešno sam izvršio fine-tuning DistilBERT modela za NER zadatak, postizući izuzetan F1 score od preko 98%. Ovi rezultati jasno demonstriraju efikasnost transfer learning pristupa u domenu obrade prirodnog jezika. Takođe, potvrđuju da "laksi" modeli poput DistilBERT-a mogu postići performanse uporedive sa većim modelima, uz pravilno podešavanje hiperparametara.

Ovaj eksperiment ne samo da je rezultirao visoko-performantnim NER modelom, već je pružio i vredne uvide u proces fine-tuninga transformerskih modela. Ovi uvidi mogu biti od velike koristi za buduće projekte u oblasti obrade prirodnog jezika, posebno kada su resursi ograničeni ili je potrebna brza adaptacija na specifične domene.

U budućim istraživanjima, bilo bi zanimljivo istražiti performanse ovog modela na drugim NER datasetovima ili srodnim zadacima, kao i uporediti ga sa većim modelima u smislu trade-offa između performansi i računarske efikasnosti.