

# TRANSLATION APP

[GitHub repository](#)



**Nikola Mitrović SV18/2021, Bojan Živanić SV61/2021**

# UVOD

- Cilj našeg projekta je razvoj aplikacije za prevođenje teksta koja koristi tehnike veštačke inteligencije.
- Aplikacija je dizajnirana da olakša komunikaciju među ljudima koji govore različite jezike, omogućavajući brzu i efikasnu translaciju teksta.
- Problem koji rješavamo je smanjenje jezičkih barijera koje često otežavaju putovanja, poslovne sastanke, obrazovanje i međunarodnu saradnju.

# MODEL


Bazira se na LSTM arhitekturi. Glavne komponente modela su:

1. Embedding sloj - transformiše ulazne riječi u vektorske reprezentacije.
2. LSTM sloj - obrađuje sekvencijalne podatke, omogućavajući modelu da zadrži informacije o dugoročnim zavisnostima u tekstu.
3. Linearni sloj - pretvara skrivene reprezentacije u vektore koji predstavljaju vjerovatnoće za svaku riječ u vokabularu.
4. Dropout sloj - koristi se za smanjenje prenaučenosti modela.


# TRANSLATION MODEL

---


```
import torch.nn as nn
```

10 usages     Nikola Mitrovic

```
class TranslationModel(nn.Module):
```

 Nikola Mitrovic

```
def __init__(self, input_size, embedding_size, hidden_size, output_size):
    super(TranslationModel, self).__init__()
    self.embedding = nn.Embedding(input_size, embedding_size)
    self.rnn = nn.LSTM(embedding_size, hidden_size, batch_first=True)
    self.fc = nn.Linear(hidden_size, output_size)
    self.dropout = nn.Dropout(0.5)
```

 Nikola Mitrovic

```
def forward(self, x):
    embedded = self.dropout(self.embedding(x))
    output, _ = self.rnn(embedded)
    output = self.fc(output)
    return output
```



# PREPROCESSING

---

- Tokenizacija teksta - tekst se dijeli na manje jedinice zvane tokeni.
- Uklanjanje nepotrebnih simbola i stop riječi
- Normalizacija teksta
- Lematizacija - proces svođenja riječi na njihov osnovni ili "korijenski" oblik.  
Npr. "trčao" bi se svelo na "trčati"
- Stemming - proces uklanjanja sufiksa iz riječi kako bi se dobio osnovni oblik riječi. Npr. "trčao" bi se svelo na "trč"

# TRENIRANJE MODELA

- **Inicijalizacija** - definišemo model, optimizator (Adam) i funkciju gubitka (CrossEntropyLoss).
- **Petlja kroz epohe** - iteriramo kroz cijeli skup podataka u više epohe.
- **Optimizacija parametara** - za svaki batch podataka:
  - Resetujemo gradijente optimizatora.
  - Prosleđujemo podatke kroz model i dobijamo izlaz.
  - Izračunavamo gubitak između predviđenog i stvarnog izlaza.
  - Unazadno širimo gubitak kroz mrežu i ažuriramo težine modela koristeći optimizator.
- **Praćenje napretka:** praćenje prosječnog gubitka tokom svake epohe kako bismo evaluirali performanse modela.

# REZULTATI TRENIRANJA

---

```
Epoch 1/75, Loss: 8.225131728432395  
Epoch 2/75, Loss: 8.000320275624594  
Epoch 3/75, Loss: 7.520933237942782  
Epoch 4/75, Loss: 6.854564695647269  
Epoch 5/75, Loss: 6.336418657591849
```

●  
●  
●  
●

```
Epoch 71/75, Loss: 0.08150070815375357  
Epoch 72/75, Loss: 0.08106727593324402  
Epoch 73/75, Loss: 0.07567879828539761  
Epoch 74/75, Loss: 0.07001009143211624  
Epoch 75/75, Loss: 0.0776365323161537
```

# EVALUACIJA

- Model se prebacuje u evaluacioni režim koristeći **model.eval()**.
- Računanje BLEU Skora:
  - Za svaki par izvornog i ciljnog teksta u dataloader-u, model generiše predviđeni tekst.
  - Kvalitet prevođenja se evaluira korišćenjem BLEU skora, koji mjeri sličnost između generisanog i referentnog teksta.
- Svaki izračunati BLEU skor se dodaje u listu scores kao mjera performansi modela za datu epohu evaluacije.
- Konačni rezultat evaluacije je prosječna vrijednost svih BLEU skorova iz liste scores, pružajući pregled ukupnog kvaliteta prevođenja modela.



# REZULTATI EVALUACIJE

---

```
BLEU score for en_sr: 0.0956
```

```
BLEU score for sr_en: 0.0396
```

# APLIKACIJA

Text Translation Application

Translation Direction: en\_to\_sr

Source Text: Me loved you!

Translated Text: ja voleo vi!

Translate

Text Translation Application

Translation Direction: sr\_to\_en

Source Text: Mama i tata su super! Ja takodje!

Translated Text: mum and dad they grea! me to!

Translate

# BUDUĆI PLANOVI

---

- Proširivanje skupa podataka
- Optimizacija arhitekture modela
- Poboljšanje razumijevanja vremena i lica
- Poboljšanje performansi
- Korišćenje dodatnih metrika evaluacije