

**Generisanje
teksta
(The Office)**



O skupu podataka

- Skup sadrži 58721 replika
- Svaka instanca sadrži:
 - ime lika
 - tekst replike
 - broj sezone
 - broj epizode
- Michael - lik sa najviše replika
- Yeah - najčešća replika



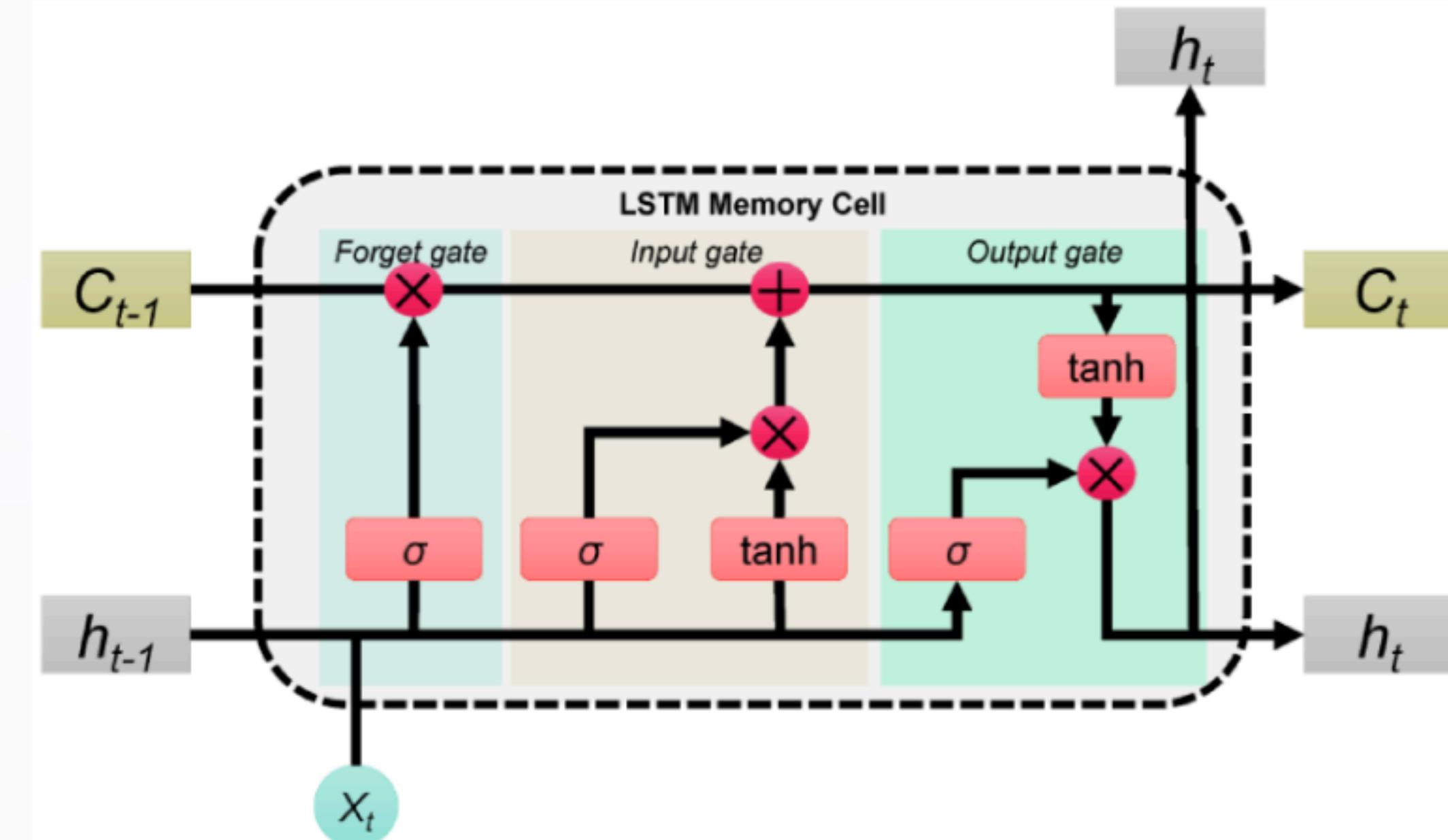
Pretprocesiranje podataka

- Ukupno ima preko 700 likova
- Izdvajanje 20 najčešćih likova - 86.3% replika
- Predugački dijalozi i one-lineri
- Podela na train i test replike



Neuronske mreže

- LSTM





LSTM modeli

word-level

no
why
me
dwight

char-level

a, c.
b, x
t

Word-level LSTM model

- Klasa Tokenizer
- Arhitektura slojeva

```
opt = Adam(learning_rate=0.0005)

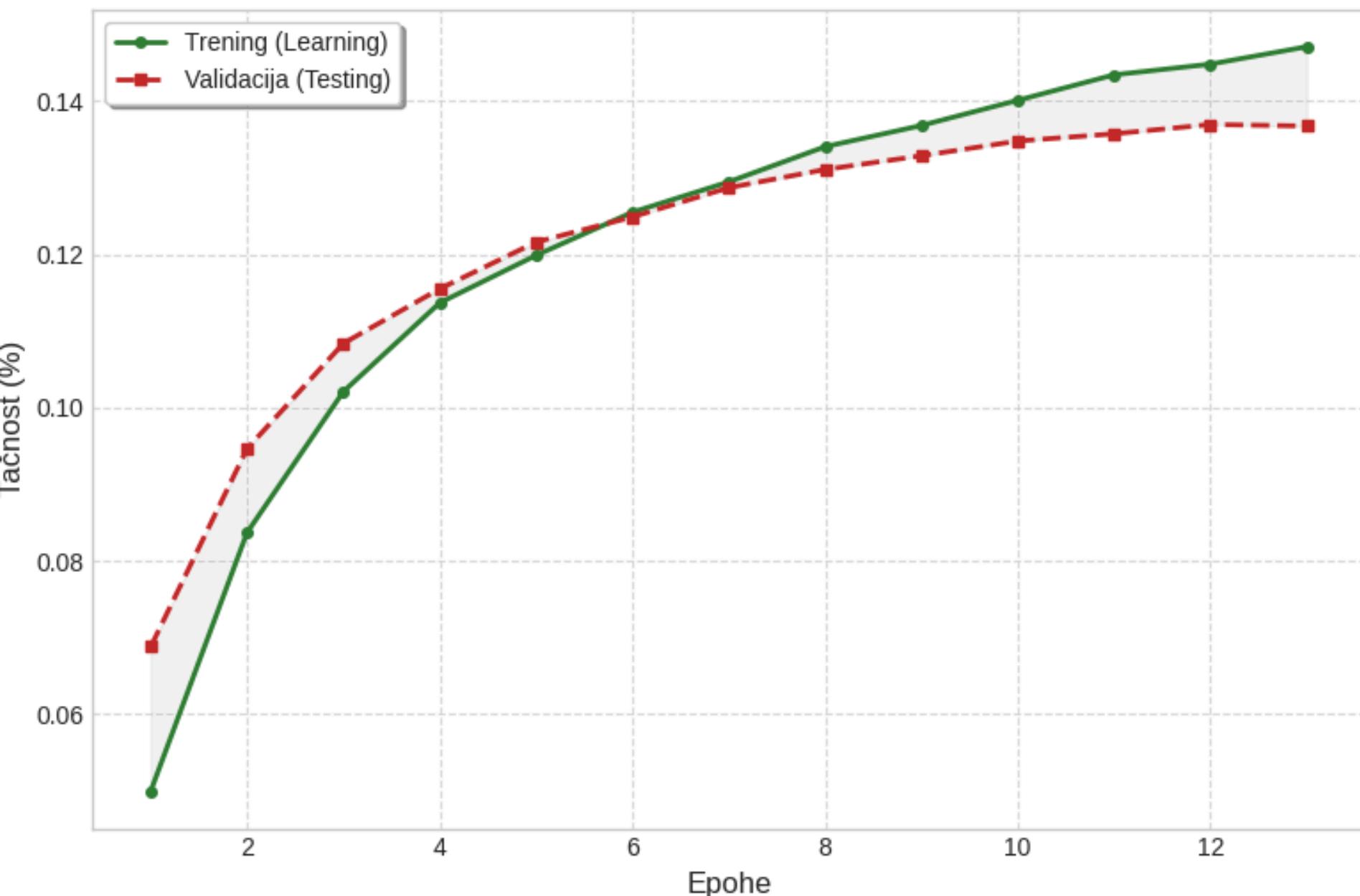
model = Sequential([
    Embedding(total_words, 150, input_length=max_sequence_len-1),
    LSTM(256, return_sequences=True),
    Dropout(0.4),
    LSTM(128),
    Dropout(0.3),
    Dense(total_words, activation='softmax')
], name="Triple_LSTM_v1_256-128")

model.compile(loss='sparse_categorical_crossentropy', optimizer=opt, metrics=['accuracy', perplexity])
model.summary()
```

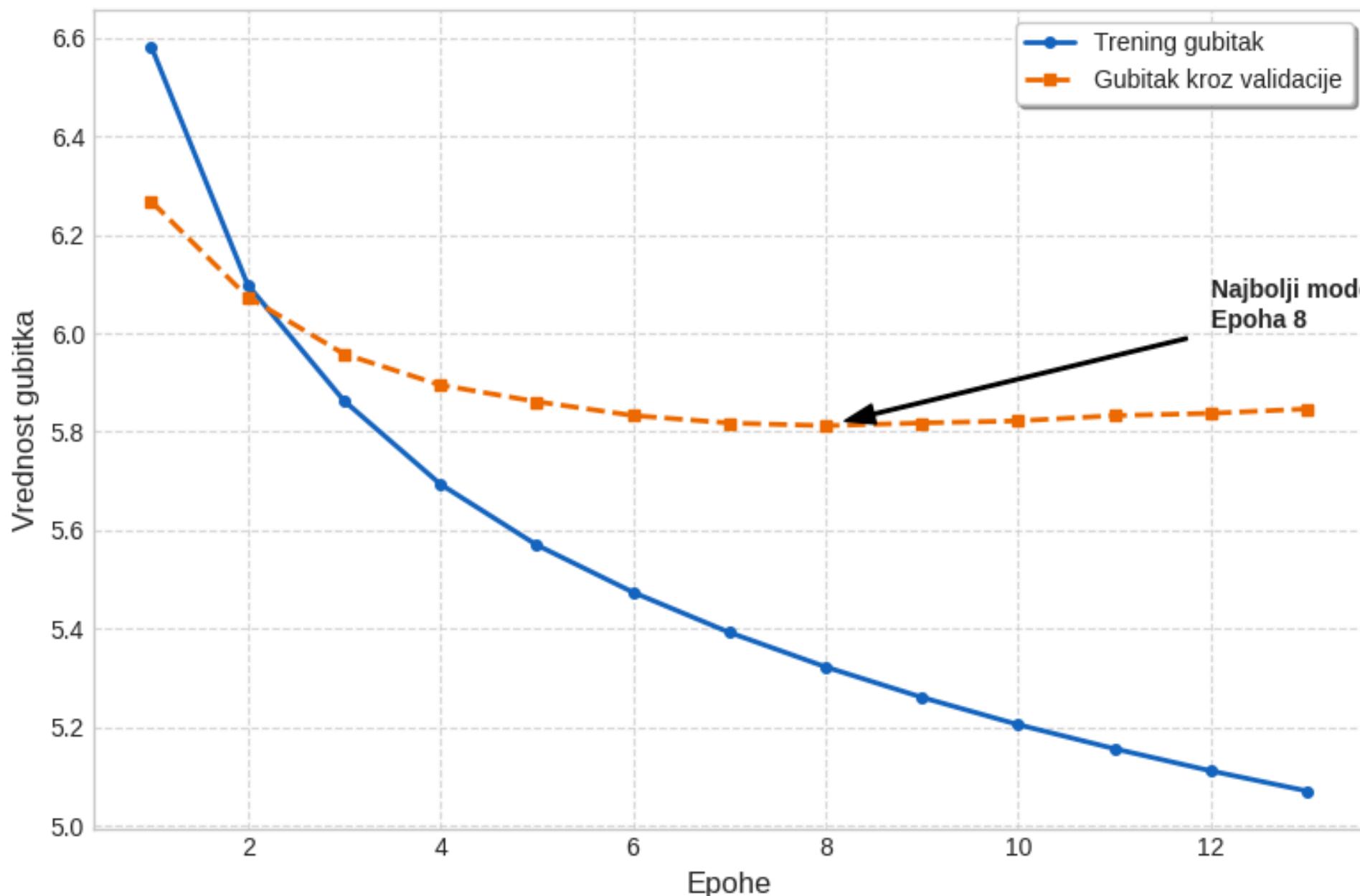
Word-level LSTM model

Analiza treninga: The Office LSTM word-level Model

Tačnost kroz epohe



Gubitak kroz epohe



Char-level LSTM model

- Mapiranje
- Sekvence

```
▶ embedding_dim = 50

lstm_units_1 = 128
lstm_units_2 = 64

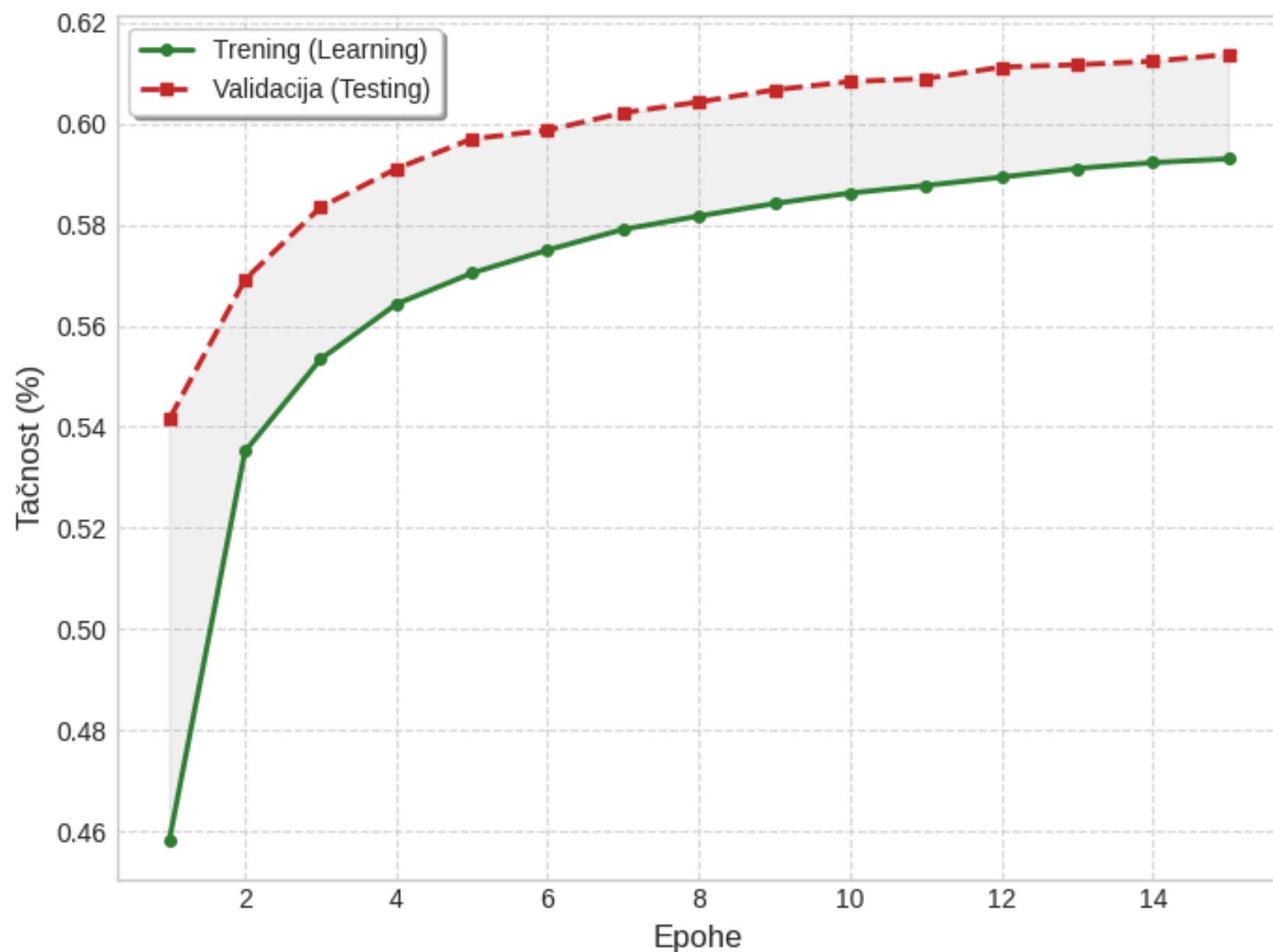
model = Sequential([
    Embedding(input_dim=vocab_size, output_dim=embedding_dim),
    LSTM(lstm_units_1, return_sequences=True),
    Dropout(0.2),
    LSTM(lstm_units_2),
    Dropout(0.2),
    Dense(vocab_size, activation='softmax')
])
```

```
model.compile(
    loss='sparse_categorical_crossentropy',
    optimizer='adam',
    metrics=['accuracy'])
```

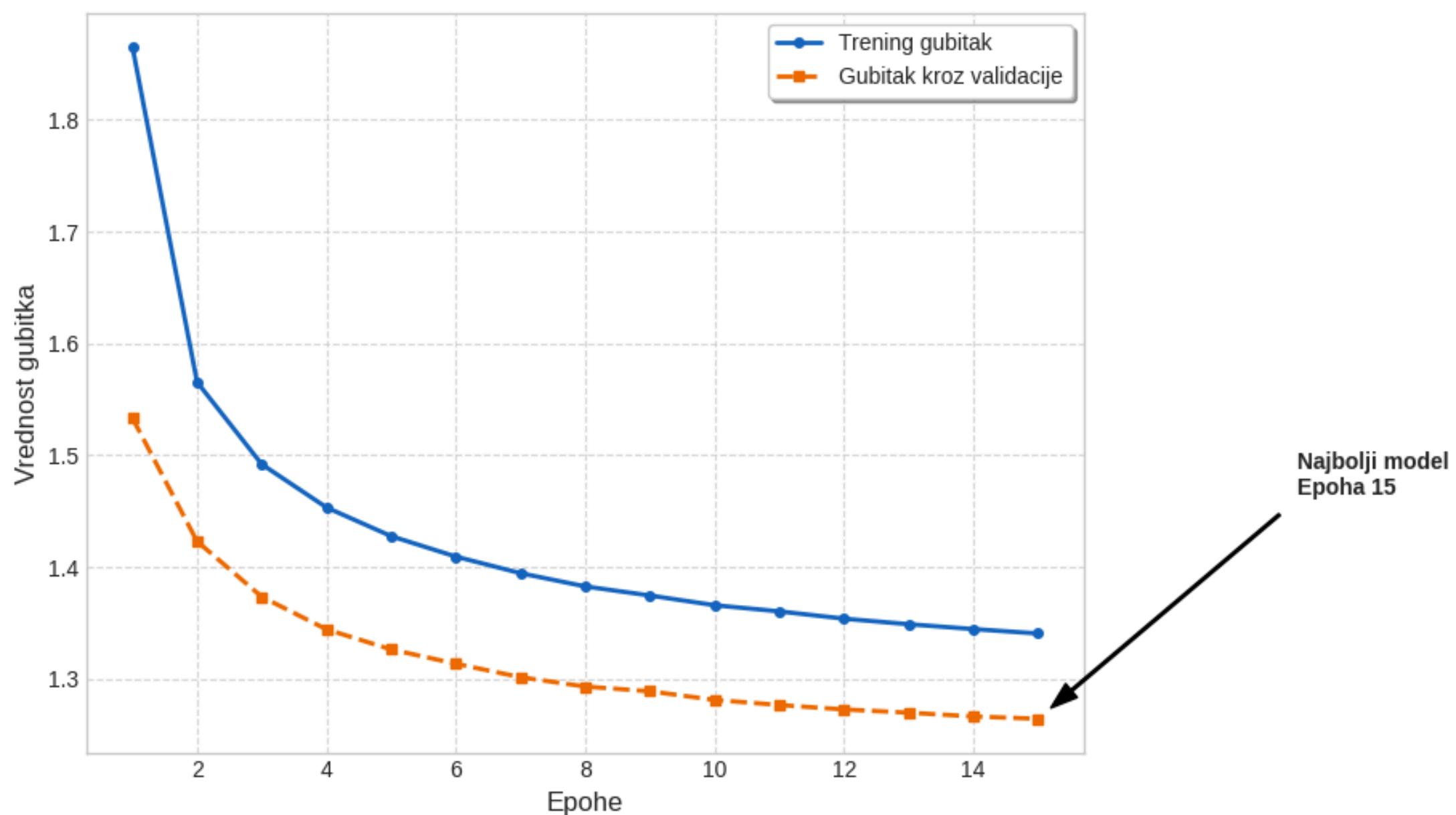
Char-level LSTM model

Analiza treninga: The Office LSTM char-level Model

Tačnost kroz epohe



Gubitak kroz epohe



Generisanje teksta

```
seed = "jim: why"
print("CHAR-LEVEL: " + generate_single_line(char_model, seed, char_to_index, index_to_char, length=70, temperature=0.6))
print("WORD-LEVEL: " + generate_text_with_temp(seed, 15, word_model, max_sequence_len, temperature=0.6))
```

CHAR-LEVEL: jim: why are you going to take in it.

WORD-LEVEL: jim: why don't you just talk to you about you guys remember that's the other friends in

```
seed = "pam: because"
print("CHAR-LEVEL: " + generate_single_line(char_model, seed, char_to_index, index_to_char, length=70, temperature=0.7))
print("WORD-LEVEL: " + generate_text_with_temp(seed, 15, word_model, max_sequence_len, temperature=0.7))
```

CHAR-LEVEL: pam: because fun alright?

WORD-LEVEL: pam: because i saw you doing it who is a great person so... good and you look



Generisanje teksta

- Parametar Temperature (T)
- $T < 0.6$ - predvidiviji
- $T > 1$ - kreativniji
- Tražimo zlatnu sredinu



Generisanje teksta

- temperature = 0.4

thats what she say it.

char

thats what she was a very good idea i don't know what you have a little name of

word

- temperature = 1

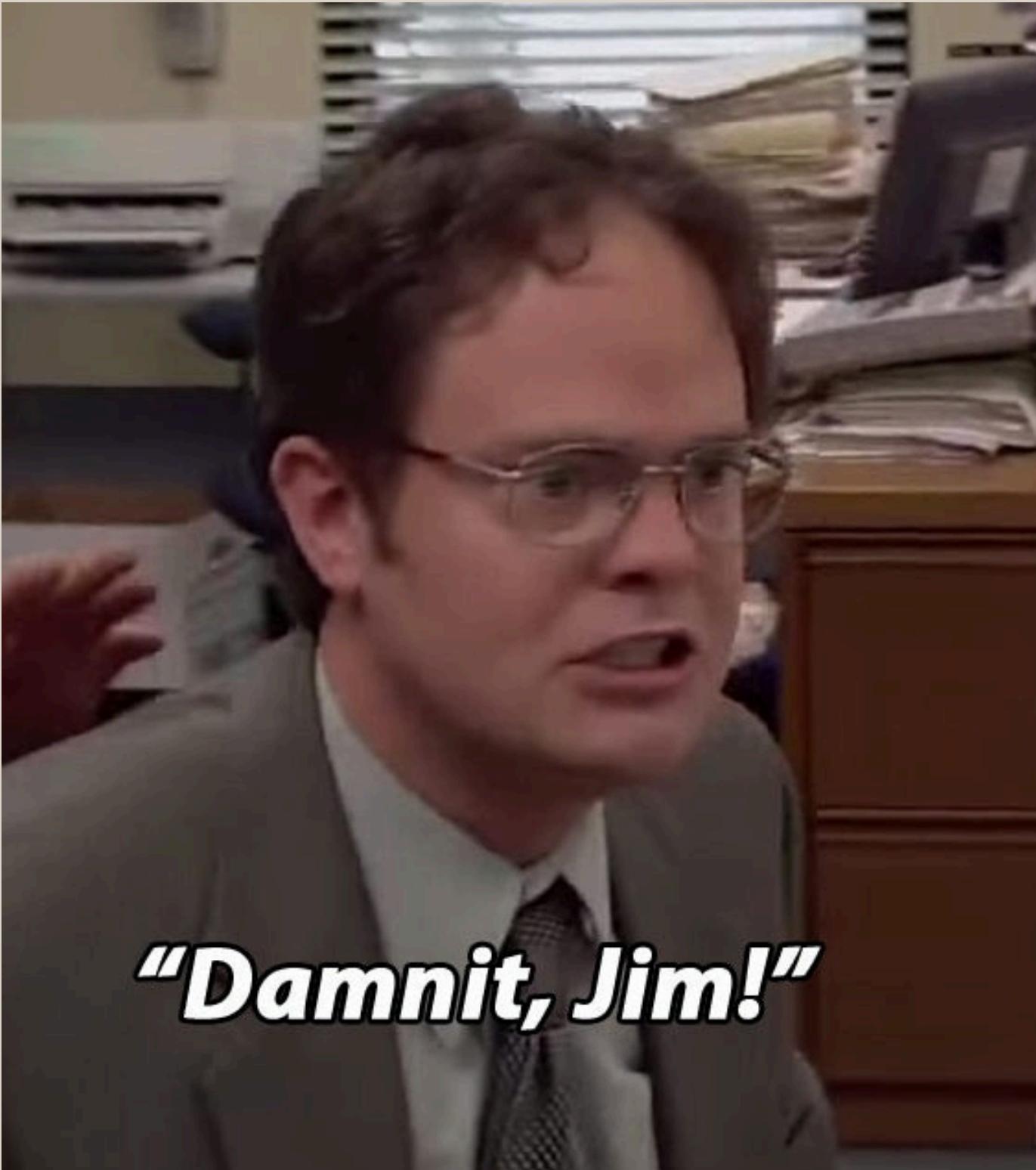
thats what she need connantly. my firtin, and thats but hes indegilf

char

thats what she water lost the office corporate do really would've will just leave her in a grill

word

Generisanje teksta



“Damnit, Jim!”



***“He put my stuff in
Jell-O again.”***

Generisanje teksta

- temperature = 0.2

dwight: he put my stuff the best of the starting the party
and the same of the company to be
dwight: he put my stuff in the office and i don't know what
you have to do that you know

char

word

- temperature = 0.9

dwight: he put my stuff. i am going to toby away deathding
affected through.

char

dwight: he put my stuff who is being to ask that... of the odd
room paper today jim kinda here

word

Zaključak

- LSTM vs Transformer
- Poređenje sa GPT-2 modelom

```
▶ generate_comparison("MICHAEL: Dwight, you are")
generate_comparison("JIM: (to camera) Yesterday,")

...
--- SEED: MICHAEL: Dwight, you are ---
LSTM: MICHAEL: Dwight, you are going to be a real good time and i got to spend the way so
GPT-2: MICHAEL: Dwight, you are a suspect.
-----
--- SEED: JIM: (to camera) Yesterday, ---
LSTM: JIM: (to camera) Yesterday, we're gonna be here of the regional phone i don't know you know what you
GPT-2: JIM: (to camera) Yesterday, I was just back from my first day at work, so I figured I'd catch you walking to
```

kraj :)

