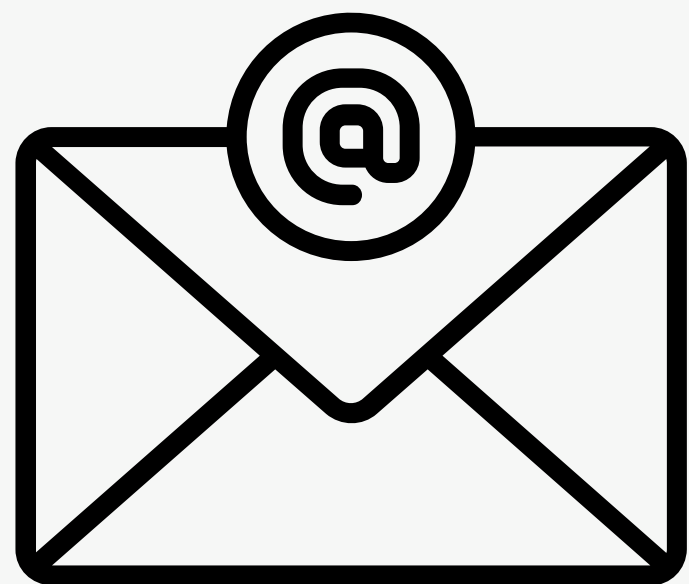


Klasifikacija spam mejlova



Katarina Gavrilović i Pavle Ivanović

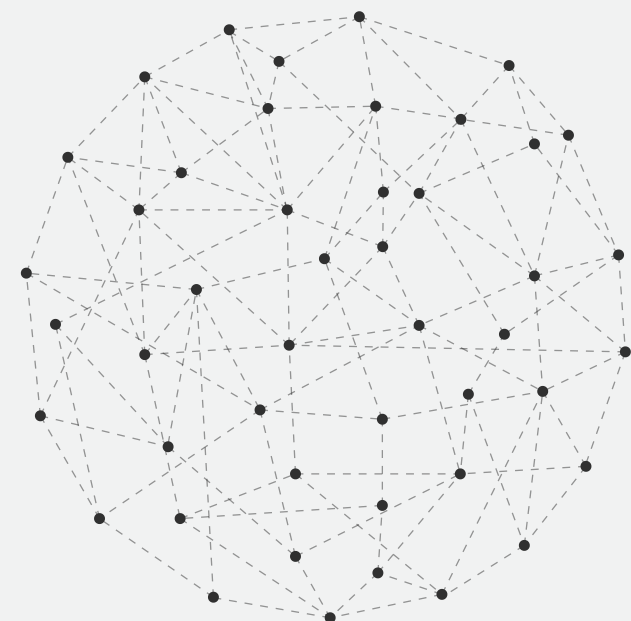
Sadržaj

01 Prikupljanje i obrada podataka

02 Naivni Bajesov klasifikator

03 Neuralna mreža

04 Rezultati



Početni set podataka

- 747 spam i 4,827 ham mejlova

	Label	SMS
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

Obrada podataka

- **Enkodovanje labela**

- **Tokenizacija**

- **Podela**

80/10/10

	Label	SMS
0	0	go jurong point crazi avail bugi n great world...
1	0	ok lar joke wif u oni
2	1	free entri wkli comp win fa cup final tkts st ...
3	0	u dun say earli hor u c already say
4	0	nah dont think goe usf live around though

Tabela1: sredjeni podaci

Naivni Bajesov klasifikator

Naivni Bajesov klasifikator

Računanje verovatnoća

$$\mathcal{P}(\text{SPAM} | \underbrace{a, b, c \dots}_{\text{tokeni}}) = \frac{\mathcal{P}(a|\text{SPAM}) \mathcal{P}(b|\text{SPAM}) \dots \mathcal{P}(\text{SPAM})}{\mathcal{P}(a) \mathcal{P}(b) \dots}$$

$$\mathcal{P}(\text{ham} | a, b, c \dots) = \frac{\mathcal{P}(a|\text{ham}) \mathcal{P}(b|\text{ham}) \dots \mathcal{P}(\text{ham})}{\mathcal{P}(a) \mathcal{P}(b) \dots}$$

$$\mathcal{P}(\text{SPAM} | a, b, c \dots) > \mathcal{P}(\text{ham} | a, b, c \dots) \Rightarrow \text{SPAM}$$

$$\mathcal{P}(\text{SPAM} | a, b, c \dots) \leq \mathcal{P}(\text{ham} | a, b, c \dots) \Rightarrow \text{HAM}$$

Neuralna mreža

Neuralna mreža

TFIDF metoda

svakom tokenu u poruci se dodeljuje brojna vrednost

$$TF = 0.5 + 0.5 \cdot \frac{\text{frekvencija tokena u poruci}}{\text{frekvencija tokena u vokabularu}}$$

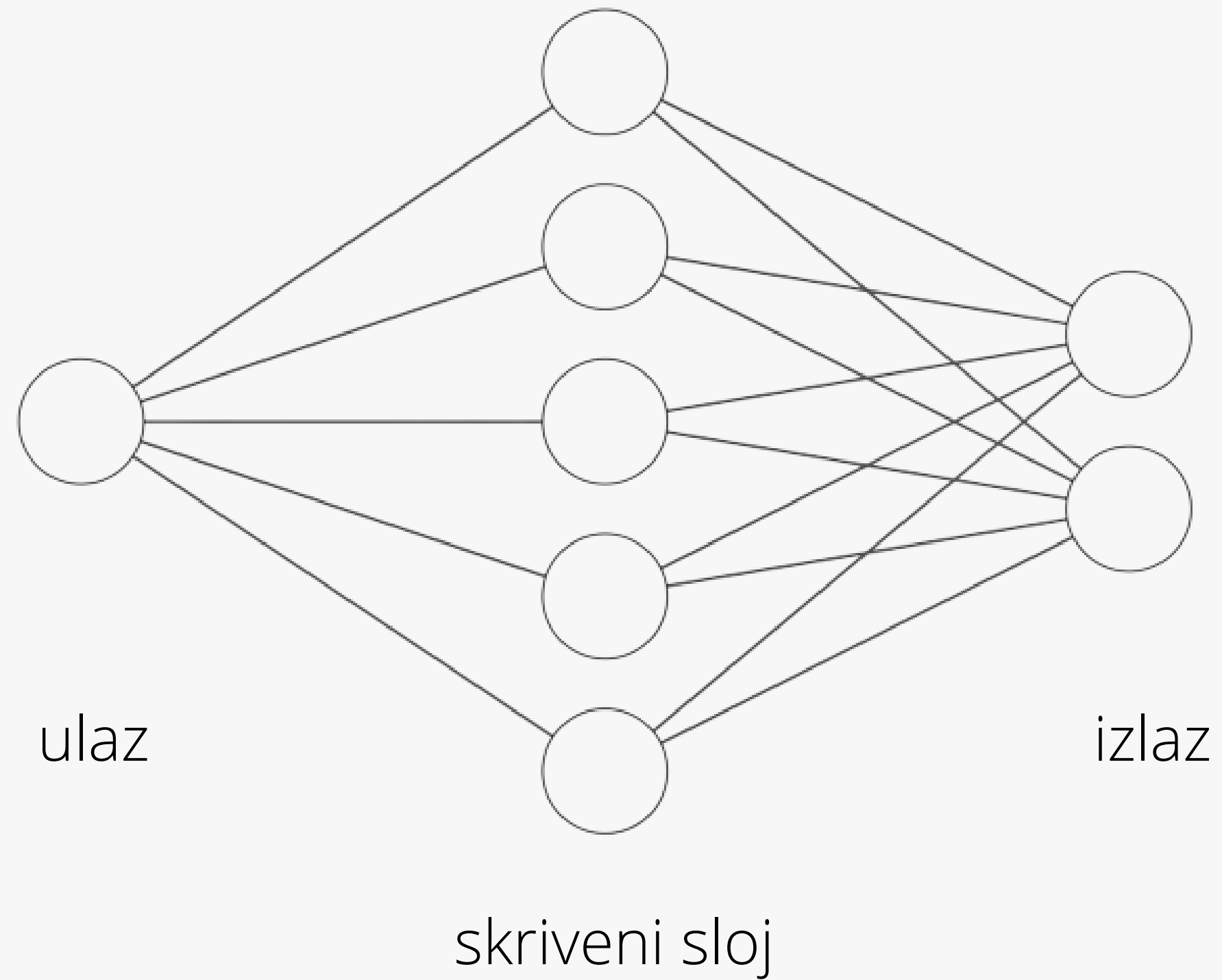
$$IDF = \log \frac{\text{ukupan broj poruka}}{\text{max frekvencija u poruci}}$$

$$TFIDF = TF \cdot IDF$$

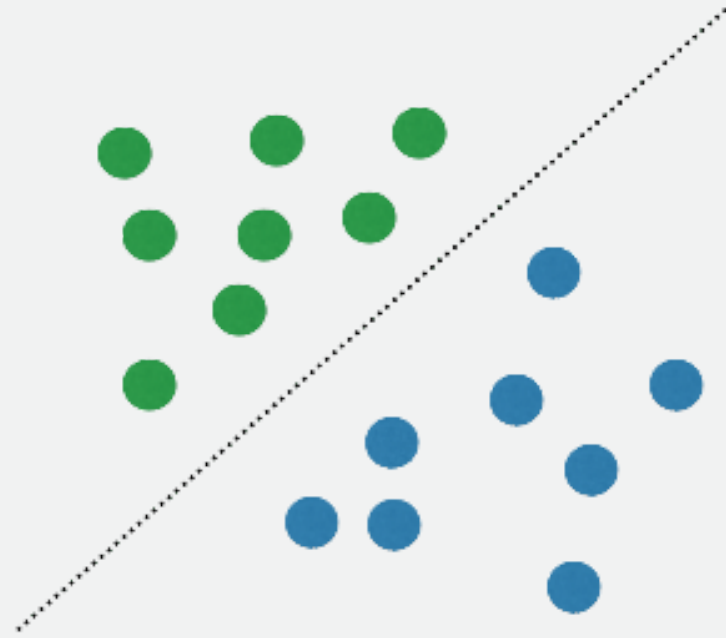
	call	ok	im	go	get	come	ill	ur	ltgt	dont	...
SMS0	0.000000	0.000000	0.0	0.140630	0.0	0.0	0.0	0.0	0.0	0.000000	...
SMS1	0.000000	0.282501	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...
SMS2	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...
SMS3	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...
SMS4	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.241897	...

Tabela2: podaci za neuralnu mrežu

Arhitektura neuralne mreže



Rezultati



Naivni Bayes

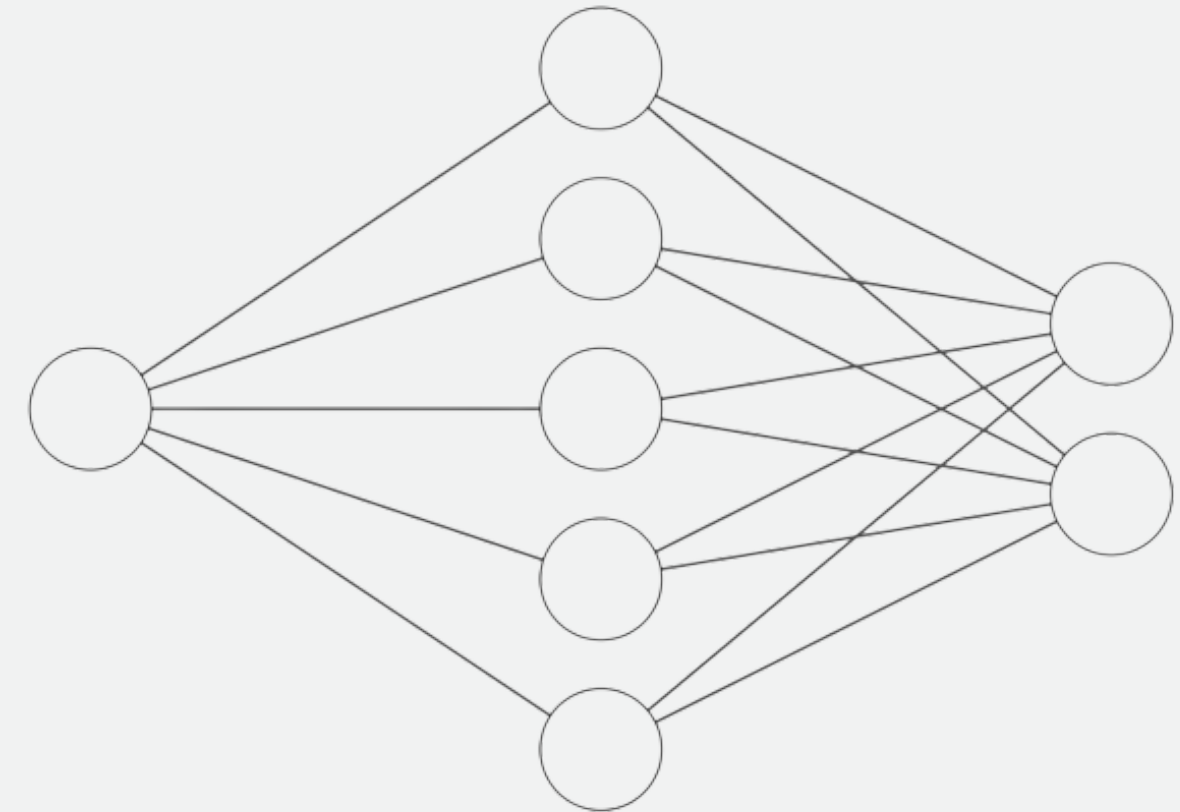
tačnost : 95.42%

preciznost: 95.19%

odziv: 68.27%

f1: 79.52%

f2: 88.24%



Neuralna mreža

tačnost : 97.76%

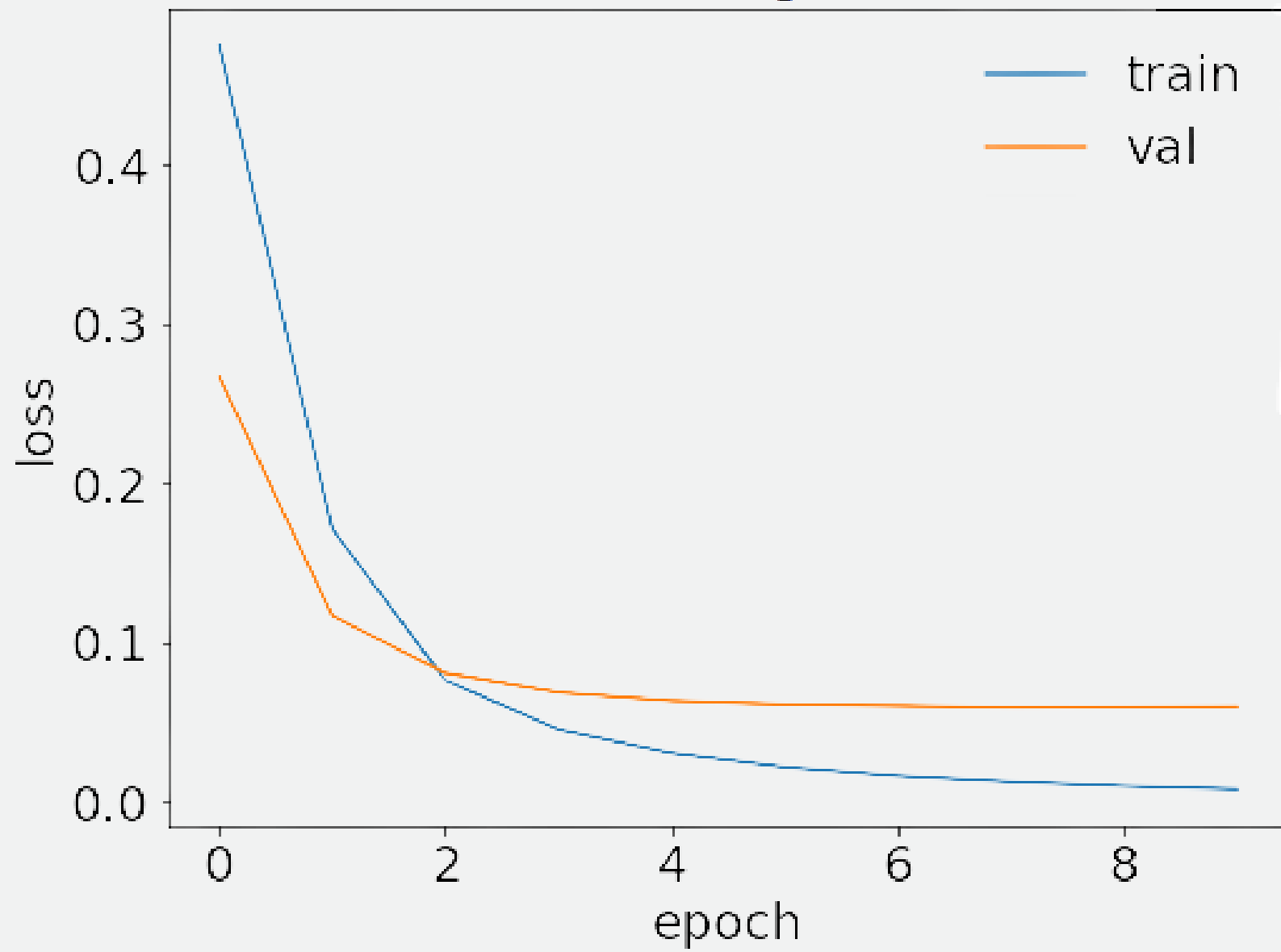
preciznost: 95.42%

odziv: 86.21%

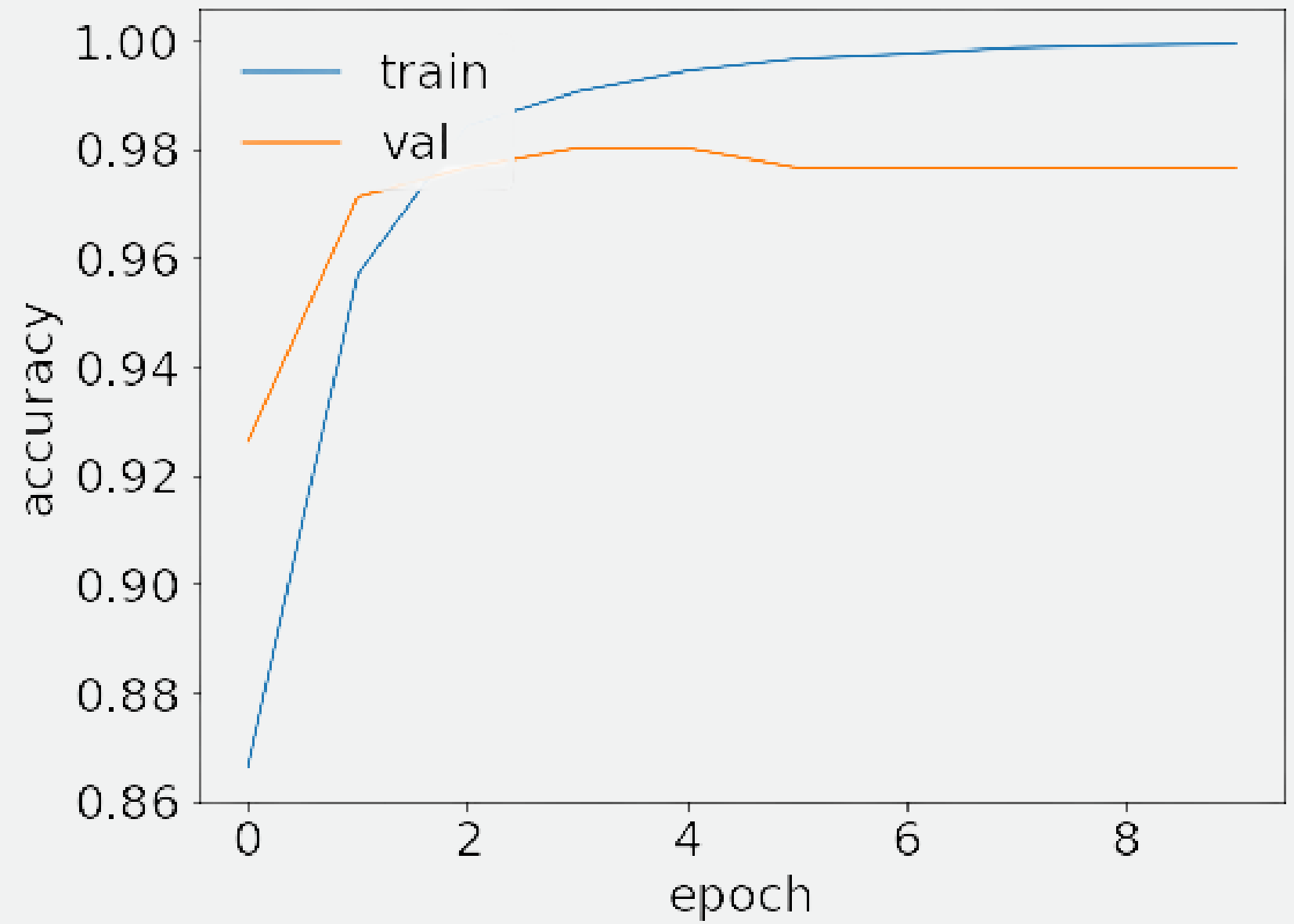
f1: 90.58%

f2: 93.42%

Grafik greške



Grafik tačnosti



Reference

Python biblioteke:

- pandas
- numpy
- matplotlib
- nltk
- sklearn
- keras

Dataset:

- [SMSSpamCollection](#)

Hvala na pažnji!

Ali nije kraj...

REF: - INSTRUCTION TO CREDIT YOUR ACCOUNT WITH THE SUM OF (US\$15M)

John Rob <johnsala222@gmail.com>

to bcc: me ▾

REF: - INSTRUCTION TO CREDIT YOUR ACCOUNT WITH THE SUM OF (US\$15M)

John Rob <johnsala222@gmail.com>

to bcc: me ▾

PAŽNJA

REF: - INSTRUCTION TO CREDIT YOUR ACCOUNT WITH THE SUM OF (US\$15M)

John Rob <johnsala222@gmail.com>

to bcc: me ▾

PAŽNJA

Dragi korisniče, posle temeljne evaluacije, Nigerijska nacionalna petroleum korporacija, zajedno sa Ministarstvom finansija Federalne republike Nigerije, je odlučila da Vas kontaktira kako bi izvršili uplatu od \$15.000.000,00.

REF: - INSTRUCTION TO CREDIT YOUR ACCOUNT WITH THE SUM OF (US\$15M)

John Rob <johnsala222@gmail.com>

to bcc: me ▾

PAŽNJA

Dragi korisniče, posle temeljne evaluacije, Nigerijska nacionalna petroleum korporacija, zajedno sa Ministarstvom finansija Federalne republike Nigerije, je odlučila da Vas kontaktira kako bi izvršili uplatu od \$15.000.000,00.

Jedino što treba da uradite jeste da nam pošaljete \$200,00 na našu BTC adresu, kako bi pokrili troškove uplate. Odgovorite nam na mejl što pre kako bi Vam rekli gde možete poslati novac. Hvala na saradnji i nadamo se da ćemo se čuti sa Vama uskoro.

REF: - INSTRUCTION TO CREDIT YOUR ACCOUNT WITH THE SUM OF (US\$15M)

John Rob <johnsala222@gmail.com>

to bcc: me ▾

PAŽNJA

Dragi korisniče, posle temeljne evaluacije, Nigerijska nacionalna petroleum korporacija, zajedno sa Ministarstvom finansija Federalne republike Nigerije, je odlučila da Vas kontaktira kako bi izvršili uplatu od \$15.000.000,00.

Jedino što treba da uradite jeste da nam pošaljete \$200,00 na našu BTC adresu, kako bi pokrili troškove uplate. Odgovorite nam na mejl što pre kako bi Vam rekli gde možete poslati novac. Hvala na saradnji i nadamo se da ćemo se čuti sa Vama uskoro.

Srdačan pozdrav,

Mr. Džon Rob

Barclays Bank Plc