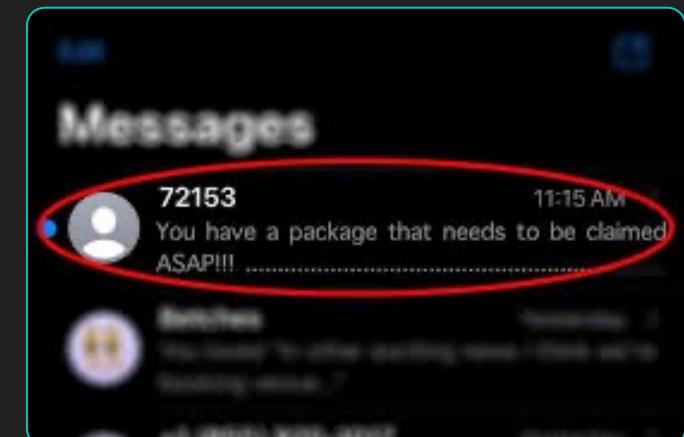
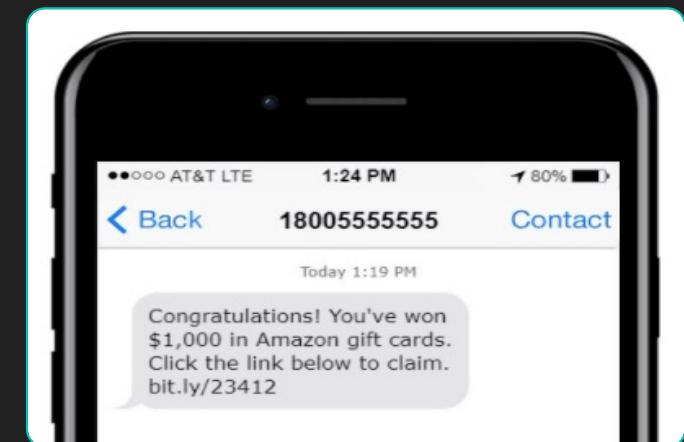


Detekcija SPAM poruka pomoću LSTM-a

Maksim Kos, Hrvoje Ljubas, Tomislav Matić, Jerko Šegvić, Lana Tuković

Detekcija SPAM poruka

- Proces identifikacije i filtriranja neželjenih poruka
- Razdvajanje legitimnih poruka od onih koje korisnici ne žele ili koje predstavljaju potencijalnu prijetnju
- Klasične karakteristike SPAM poruka: nepoželjna promocija, prijevare, neprimjeren sadržaj, lažne nagrade

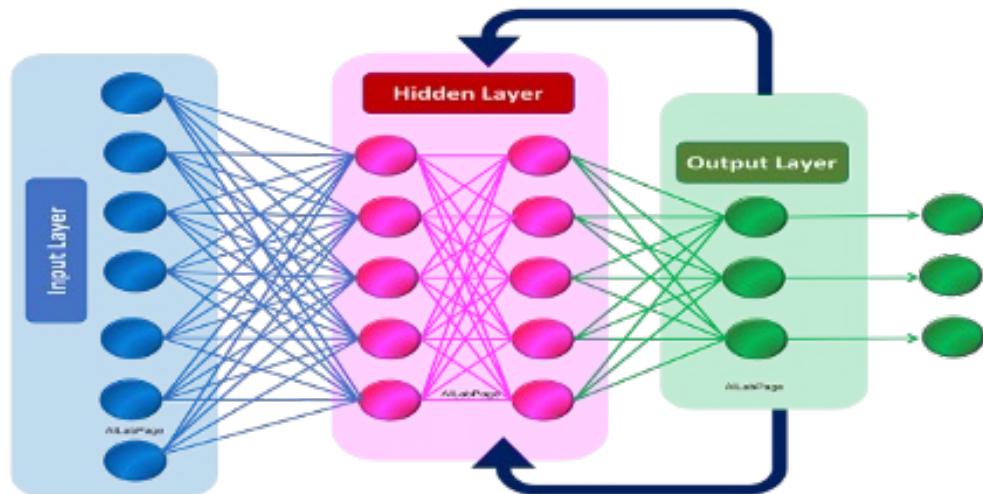


LSTM

- *Long short-term memory*
- Poseban tip rekurentne neuronske mreže (RNN) dizajniran za obradu i modeliranje sekvencijskih podataka, kao što su tekstovi u e-mailovima ili SMS porukama

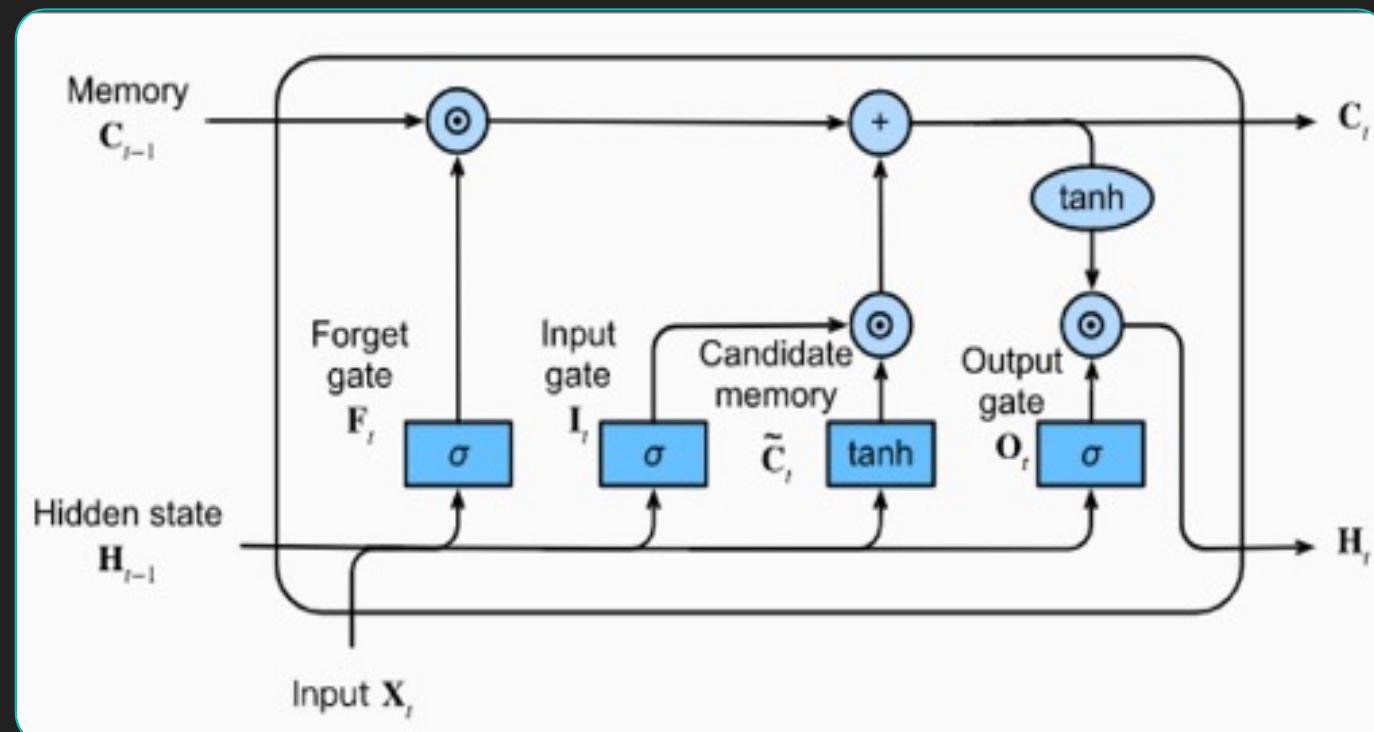
- Za razliku od klasične neuronske mreže, RNN imaju povratnu petlju, što im omogućuje održavanje unutarnjeg stanja ili "memorije"

Recurrent Neural Networks



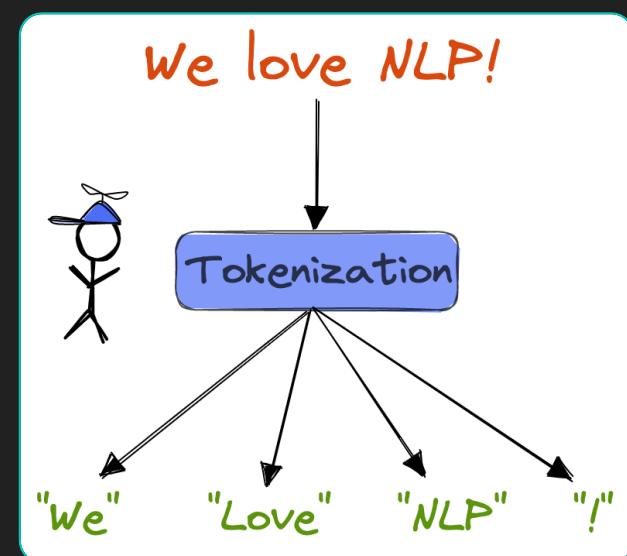
LSTM

- Standardne RNN-ove karakterizira problem nestajućeg i eksplodirajućeg gradijenta, što ograničava njihovu sposobnost održavanja dugoročnih zavisnosti
- LSTM rješava taj problem uvođenjem posebnih mehanizama poput ćelija s vrata i gate-ova



PRIPREMA PODATAKA

- Tokeniziranje teksta -> omogućava računalu da bolje razumije i interpretira tekst na temelju značajki svakog pojedinog tokena
- Tokeni su osnovne građevne jedinice kojima se tekst razdvaja, a mogu predstavljati riječi, znakove interpunkcije, brojeve ili druge segmente teksta
- Indeksiranje -> tekst možemo pretvoriti u niz indeksa pripadajućeg tokena (svakom tokenu dodijelimo indeks)
- Kraće nizove indeksa je potrebno nadopuniti tako da svi nizovi budu jednakе duljine



ARHITEKTURA MREŽE

Embedding

LSTM

Dropout

Dense

Dropout

Dense

PODEŠAVANJE HIPERPARAMETARA

- Pretraživanje po rešetci -> najbolja kombinacija hiperparametara
- Hiperparametri koje je potrebno podesiti:
 - broj memorijskih jedinica u LSTM sloju
 - vjerojatnost isključivanja neurona u Dropout slojevima
 - broj neurona u potpuno povezanim slojem
 - aktivacijske funkcije za LSTM sloj i potpuno povezani sloj

REZULTATI

- 2 LSTM mreže: jedna trenirana za SMS poruke, druga trenirana za e-mailove
- Svaka od mreža trenirana je na 5000-6000 podataka
- Optimizacijom hiperparametara mreža postignuta je točnost od 99.48% za klasifikaciju SMS-ova te 95.6% za klasifikaciju e-mailova
- Dodatan validacijski korak: mreže trenirane na SMS porukama testirane nad e-mailovima i obrnuto -> pad u performansama

ZAKLJUČAK

- Testiranje nad pripadajućim mrežama -> visoka točnost klasifikacije SMS poruka i e-mailova
- Mreža trenirana na jednom tipu poruka testirana na drugom tipu -> pad preformansi modela zbog specifičnosti strukture i sadržaja između SMS poruka i e-mailova
- Nedostatak prilagodljivosti modela ovisno o vrsti poruka koja se analizira

Hvala na pažnji!