

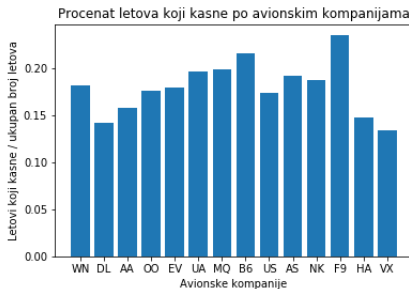
# Sinteza programa

Anja Ivanišević  
Vesna Katanić

maj 2018.

# Uvod

- ▶ Porast avionskog saobraćaja u poslednjih 20 godina
- ▶ Ogromni gubici zbog kašnjenja aviona



# Podaci

- ▶ podaci o letovima u SAD za 2015 godinu
- ▶ preko 5 miliona redova
- ▶ 30 atributa
  - ▶ podaci o letovima (day, day of the week, airline, flight number, tail number)
  - ▶ podaci o polaznom i dolaznom aerodromu (origin airport, destination airport)
  - ▶ informacije o poletanju (scheduled departure, departure time, departure delay, taxi-out, wheels-off)
  - ▶ informacije o letu (scheduled time, elapsed time, air time, distance)
  - ▶ informacije o dolasku (wheels-on, taxi-in, scheduled arrival, arrival time, arrival delay)
  - ▶ dodatne informacije (air system delay, security delay, airline delay, late aircraft delay, weather delay)

# Priprema podataka

- ▶ 300 000 random redova
- ▶ indikator da li let kasni ili ne
- ▶ jednak broj letova koji su zakasnili i onih koji nisu

# Primenjeni algoritmi

- ▶ Logistička regresija
- ▶ Potporni vektori
  - ▶ linerni
  - ▶ rbf kernel
- ▶ Neuronske mreže
  - ▶ 3 sloja
  - ▶ prvi sloj 100 neurona, drugi 40
  - ▶ na prva dva sloja *relu* aktivaciona funkcija, na poslednjem sigmoidna funkcija

# Problemi

- ▶ traženje najboljih parametara za svm sa rbf kernelom
- ▶ neuronske merže za regresiju
- ▶ bolji rezultati sa uzetih prvih 300 000 redova

# Dobijeni rezultati

Mere kvaliteta			
	Tačnost	Odziv	F1 mera
Logistička regresija	0.910	0.889	0.907
Linearni svm	0.909	0.889	0.907
Rbf	0.904	0.883	0.902
Neuronska mreža	0.929	0.886	0.907

- ▶ link ka aplikaciji
  - ▶ [https://github.com/anja-ivanisevic/ML\\_projekat](https://github.com/anja-ivanisevic/ML_projekat)
- ▶ link ka podacima
  - ▶ <https://www.kaggle.com/usdot/flight-delays/data>
- ▶ Literatura
  - ▶ <http://cs229.stanford.edu/proj2017/final-reports/5243248.pdf>
  - ▶ <http://ml.matf.bg.ac.rs/readings/ml.pdf>



Hvala na pažnji!