

## SOFTVER ZA PREDIKCIJU SMRTI LIKOVA IZ SERIJE „IGRA PRESTOLA“ -Koliko šansi za opstanak ima vaš omiljeni lik?-

### 1. Uvod

**Igra prestola** (Game of thrones) je popularna serija zasnovana na serijalu knjiga epske fantastike „Pisma leda i vatre“ Džordža Martina. Serija je poznata po borbi mnogobrojnih likova za vlast, intrigama i velikom broju smrti. Softver koji predviđa šanse za opstanak svakog lika zasnovan je na algoritmu neuronske mreže i multivarijantne i jednostavne linearne regresije. Kako se gledaoci poistovećuju sa likovima serije i veoma teško podnose smrti nekih od likova uz pomoć ove aplikacije moći će da vide kolike šanse za opstanak u narednoj sezoni ima svaki od likova. Softver je prevashodno namenjen ljubiteljima serije i onima koji se bave mašinskim učenjem i veštačkom inteligencijom.

### 4. Regresija

Regresija se koristi kao provera međusobne zavisnosti i povezanosti ulaza sa izlazom. U slučaju data seta, korišćena je višestruka regresija. Model je predstavljen jednačinom:  $y = ax_1 + bx_2 + cx_3 + dx_4 + ex_5 + fx_6 + gx_7 + hx_8 + ix_9 + s$ , gde y predstavlja matricu izlaza, a  $x_1-x_9$  predstavljaju ulaze i takođe su matričnog formata.

$$a = \frac{n \left( \sum_{i=1}^n x_i y_i \right) - \left( \sum_{i=1}^n x_i \right) \left( \sum_{i=1}^n y_i \right)}{n \left( \sum_{i=1}^n x_i^2 \right) - \left( \sum_{i=1}^n x_i \right)^2}$$

$$b = \frac{1}{n} \left( \left( \sum_{i=1}^n y_i \right) - a \left( \sum_{i=1}^n x_i \right) \right)$$

Slika 3. – Formule za izračunavanje koeficijenata kod jednostavne lin.reg.

Za testiranje regresije na konkretnom data setu korišćena je biblioteka *Accord* sa pomoćnim funkcijama i klasama.

### 6. Matrica konfuzija

Matrica konfuzija se u polju mašinskog učenja koristi prilikom problema klasifikacije. Takođe matrica konfuzije je jedan od pokazatelja ispravnosti modela i dobijenih rezultata.

Oblik matrice:

**TP** - True positives **FN** - False negatives,  
**FP** - False positives **TN** - True negatives

	Predicted is alive	
	Yes	No
Actual is alive	Yes TP FN	
	No FP TN	

### 2. Data set

Shodno vremenu nastanka softvera (pre i u toku 6. sezone serije) podaci nad kojima je rađeno mašinsko učenje obuhvataju informacije o svim likovima koji se pojavljuju ili spominju od prve do pete knjige. Data set kojim je obučena neuronska mreža obuhvata više od 800 likova i pretežno je opisan na binarni način. Sadržaj data seta:

- S.No** - redni broj;
- Name** - ime lika;
- Male** - 0= žensko, 1= muško;
- Book1** – pojavljivanje u prvoj knjizi;
- Book2** – pojavljivanje u drugoj knjizi;
- Book3** – pojavljivanje u trećoj knjizi;
- Book4** – pojavljivanje u četvrtoj knjizi;
- Book5** – pojavljivanje u petoj knjizi;
- Is Noble** – poreklo (0= nije plemićko, 1= jeste);
- Number of dead relations** – broj mrtvih srodnika;
- Is Popular** – da li je lik popularan;
- Age** – ukoliko je lik mrtav, koliko godina je imao u trenutku smrti;
- Popularity** – popularnost(decimalna vrednost);
- Is Alive** – da li je lik živ (0= nije, 1= jeste);

### 5. Rezultati

#### 5.1 Rezultati neuronske mreže

Nakon što je neuronska mreža obučena podacima iz train skupa odrađeno je treniranje sa podacima iz test skupa koji su istog formata ali obuhvataju druge likove iz serijala. Postignuta tačnost je oko 0.72% i varira u odnosu na broj međuslojeva za minimalne vrednosti 0.01 – 0.03%.

Konkretno, za lika *Arya Stark* neuronska mreža predvidela je **13.61%** šansi za opstanak u 6. sezoni. Za lika *Jon Snow*, šanse su nešto veće i iznose **15.46%**, dok za lika *Alis Fray* šanse iznose **91.96%**. Iz prethodno navedenog se može zaključiti da su šanse popularnijih likova u seriji mnogo manje nego onih manje poznatih.

#### 5.2 Rezultati višestruke linearne regresije

Vrednosti koeficijenata su (redom): -0.06, -0.07, 0, 0.19, -0.01, -0.06, -0.06, -0.36, -0.04, 0.82. Greška koju je regresija vratila iznosi 117.82. Shodno rezultatima koji su postignuti primenom višestruke linearne regresije, može se zaključiti da problem nije linearne prirode, a takođe format ulaza koji je binarnog tipa može predstavljati problem u postavci.

$$TP\ rate = \frac{TP}{TP + FN}$$

Osetljivost modela

$$FP\ rate = \frac{FP}{FP + TN}$$

Specifičnost modela

$$Precision = \frac{TP}{TP + FP}$$

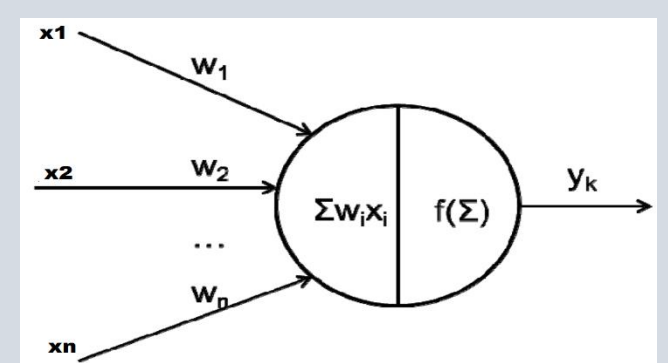
Preciznost modela

Dobijeni rezultati iznose:

**TP rate = 85% FP rate = 51% Precision=78%**

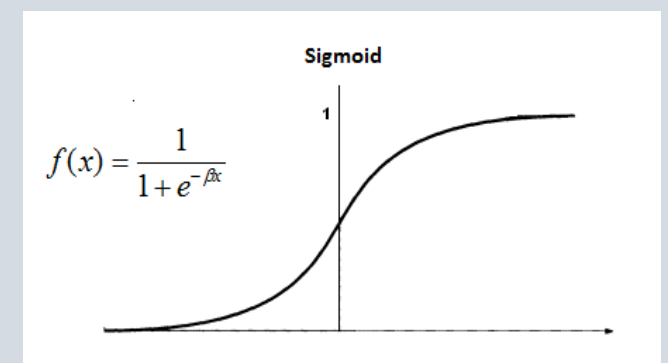
### 3. Neuronska mreža

Predviđanje šansi za opstanak likova iz data seta odrađeno je putem neuronske mreže. Mreža ima 9 ulaza iz data seta(pol, pojavljivanje u knjigama, popularnost dec., broj mrtvih srodnika, poreklo). Izlaz neuronske mreže predstavlja informacija da li je lik živ ili ne. Obučavanje neuronske mreže zasniva se na promeni težina svake od grana između neurona i primeni backpropagation algoritma. Korišćena je mreža sa jednim međuslojem.

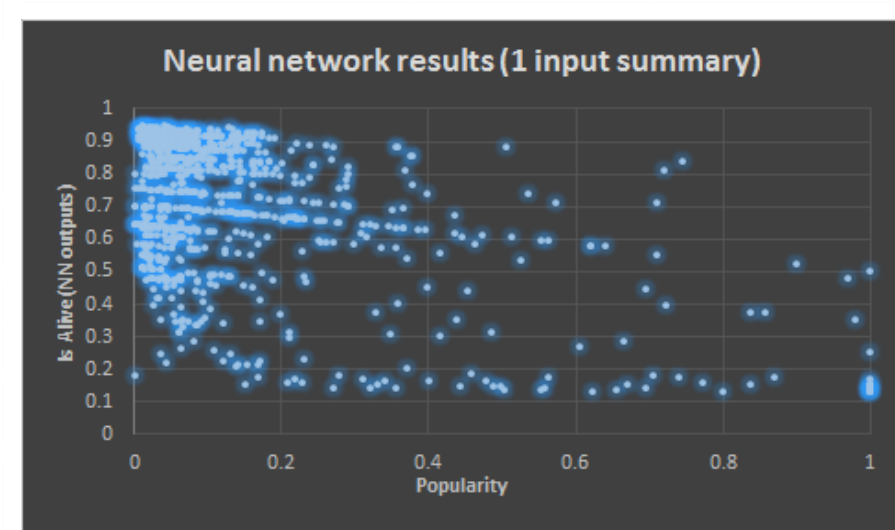


Slika 1. – Izgled neurona

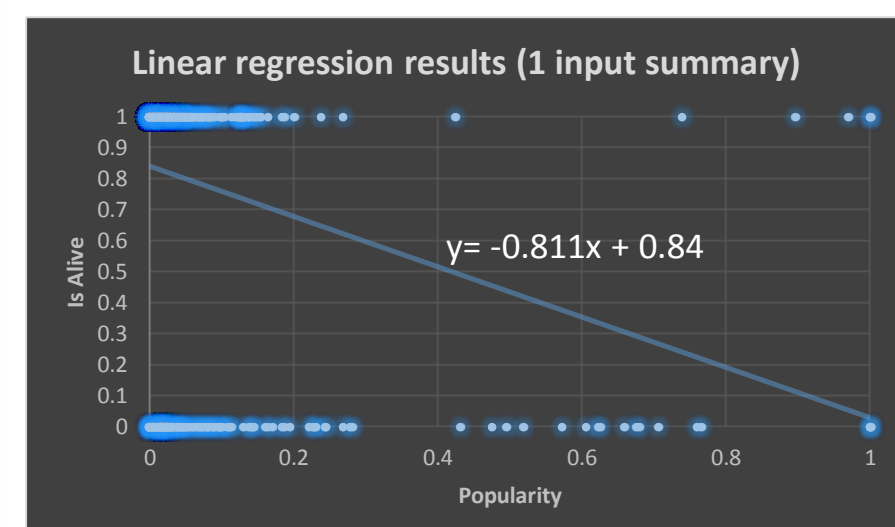
Za aktivacionu funkciju neuronske mreže uzet je sigmoid.



Slika 2.- Aktivaciona funkcija



Slika 4. – Rezultati neuronske mreže nad testnim skupom



Slika 5. – Rezultati linearne regresije nad testnim skupom