Mikuš Filip

I-SUNS – Zadanie č. 1

Úprava datasetu

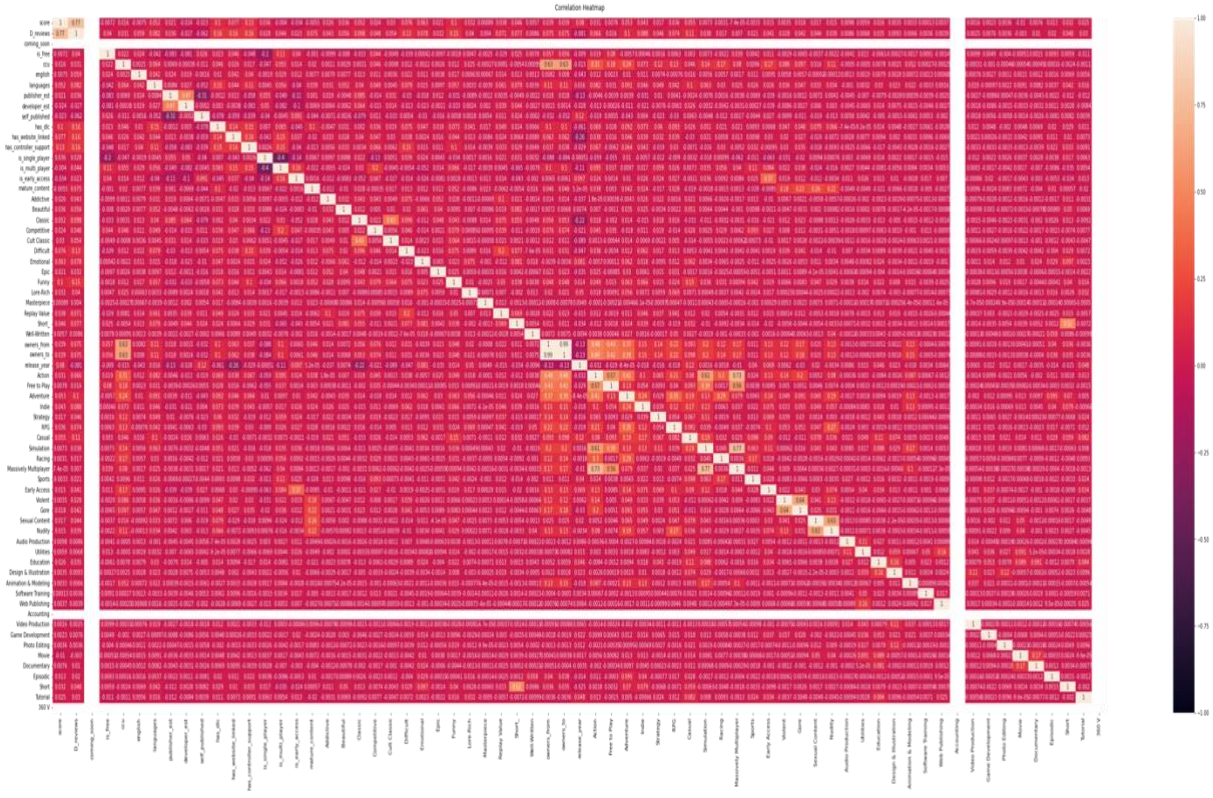
V prvom kroku sme analyzovali obsah datasetu, ktorý sme následne upravili. Konkrétne sme realizovali:

- odstránenie stĺpcov ***D_appid, D_name, VYMAZAT_price, positive, negative, D_developer, D_publisher***
- odstránenie záznamov s NaN hodnotami
- dopočítanie priemernej hodnoty skóre v záznamoch s chýbajúcou hodnotou ***score***
- nahradenie slovných recenzií v stĺpci ***D_reviews*** hodnotami z intervalu <0.0, 1.0>
- extrakcia počtu vlastníkov zo stĺpca ***D_owners*** do stĺpcov ***owners_from*** a ***owners_to***
- extrakcia roku vydania (***release_year***) z dátumu vydania (***D_release_date***)
- extrakcia žánrov (***D_genre***) a ich početností v tagoch (***D_tags***) do samostatných stĺpcov
- nahradenie hodnôt typu True/False hodnotami 0/1
- Min-Max normalizácia

EDA

Korelačná matica

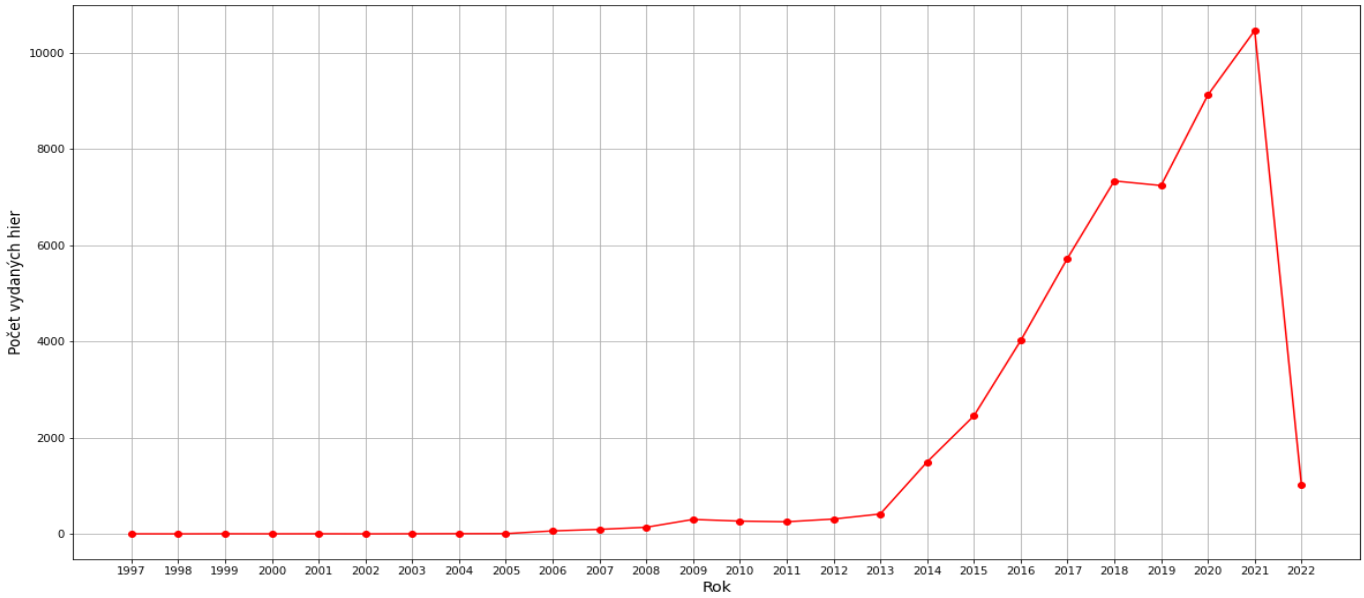
Akú majú závislosť jednotlivé dátové stĺpce?



Mikuš Filip

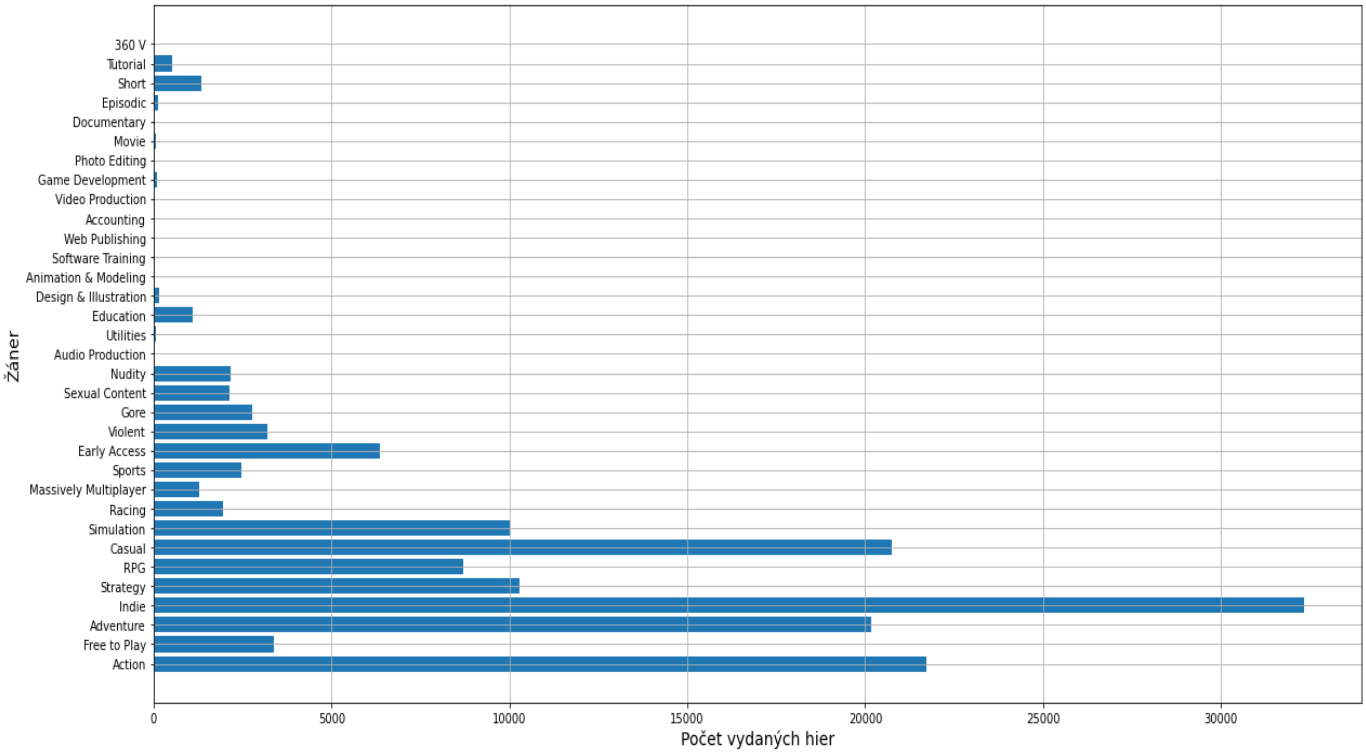
Počet hier vydaných v priebehu času

Aký je počet hier v priebehu času existencie platformy Steam? Na grafe je možné vidieť, že výraznejší rast vydávania titulov započal v roku 2014 až po súčasnosť (2022). V roku 2019 je možné pozorovať mierny pokles z dôvodu prvej vlny pandémie ochorenia COVID-19.



Počet hier jednotlivých žánrov

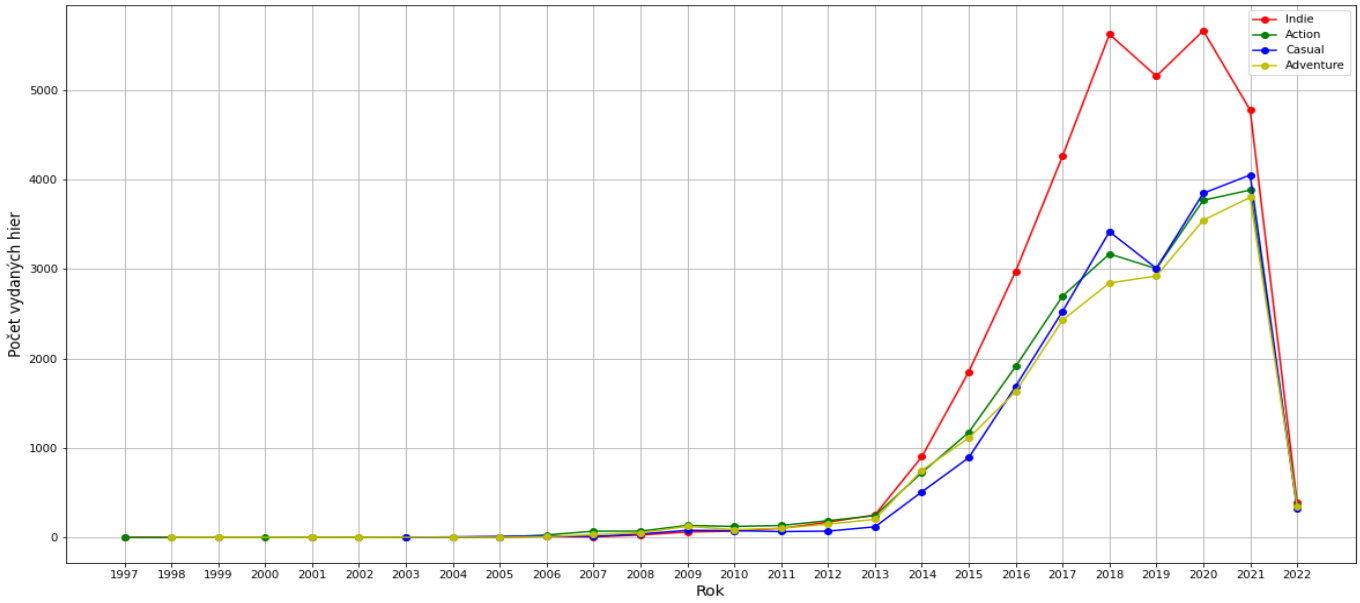
Aký je počet hier v jednotlivých žánroch? Na grafe je možné vidieť že najvyšší počet hier bol vydaný v žánroch Indie, Action a Casual.



Mikuš Filip

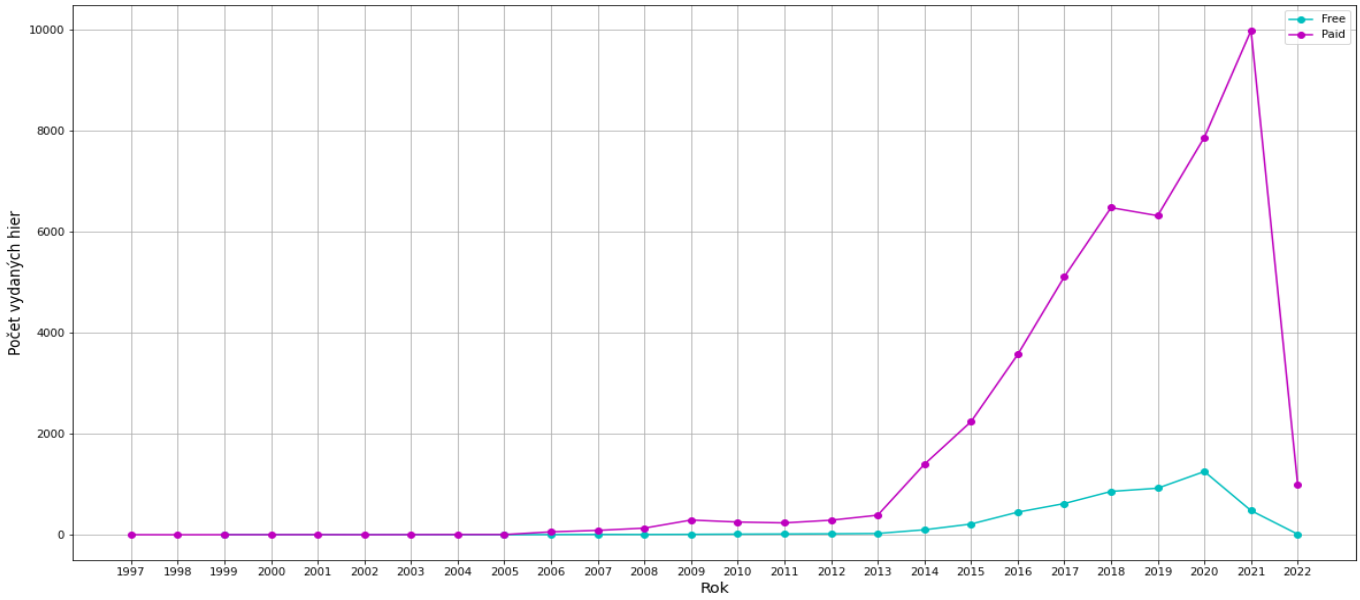
Počet hier vydaných v prebehu času v najpočetnejších žánroch

Aký bol počet hier vydaných v priebehu času v najpočetnejších žánroch? Z grafu je zrejmé. Že nastavené trendy v počtoch sú zhodné. Najvýraznejšie sú viditeľne zastúpené hry žánru Indie.



Počet hier vydaných v priebehu času podľa finančného modelu

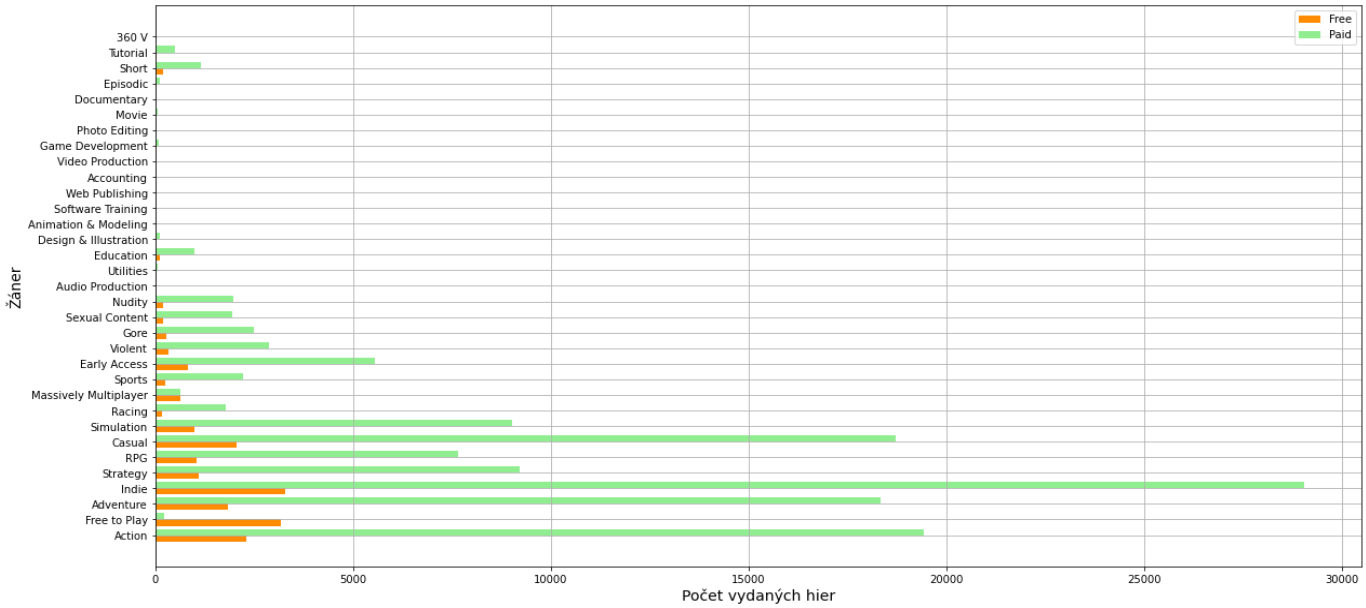
Aký bol počet hier vydaných v priebehu času podľa finančného modelu? Na grafe je možné vidieť výraznú prevahu platených titulov. Zaujímavým faktom je stúpajúci počet bezplatných titulov počas prvej vlny pandémie ochorenia COVID-19 v roku 2019.



Mikuš Filip

Počet hier jednotlivých žánrov podľa finančného modelu

Aký bol počet hier jednotlivých žánrov podľa finančného modelu? Na grafe môžeme vidieť pomer počtov hier podľa finančného modelu v jednotlivých žánroch. Zaujímavosťou je, že v žánri ***Free to Play*** sa nachádzajú aj platené tituly.



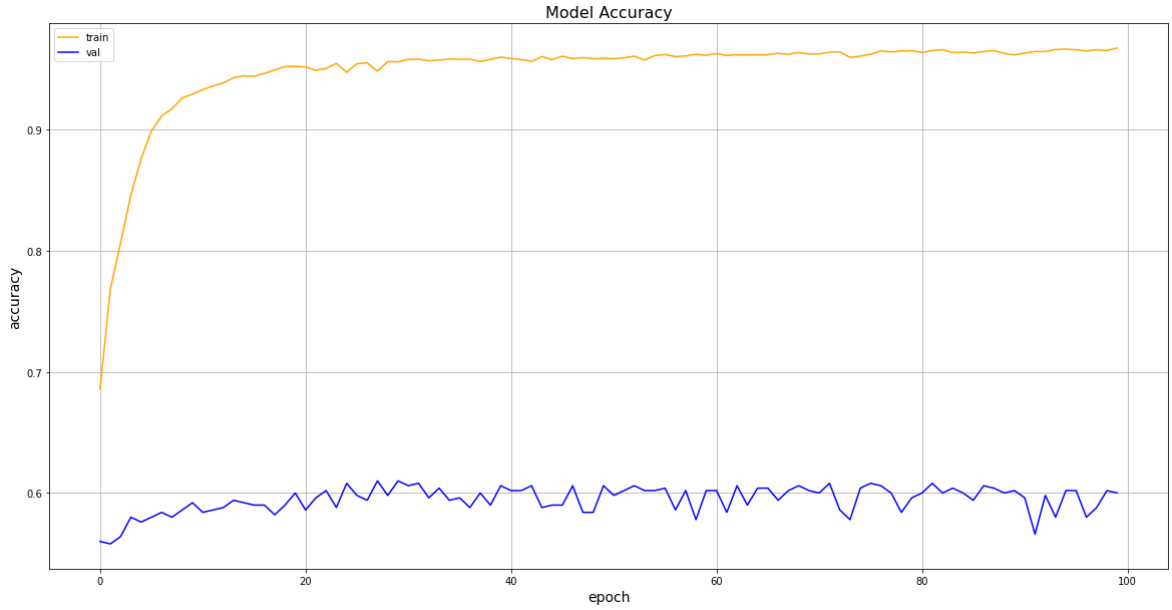
Súbor experimentov č.1

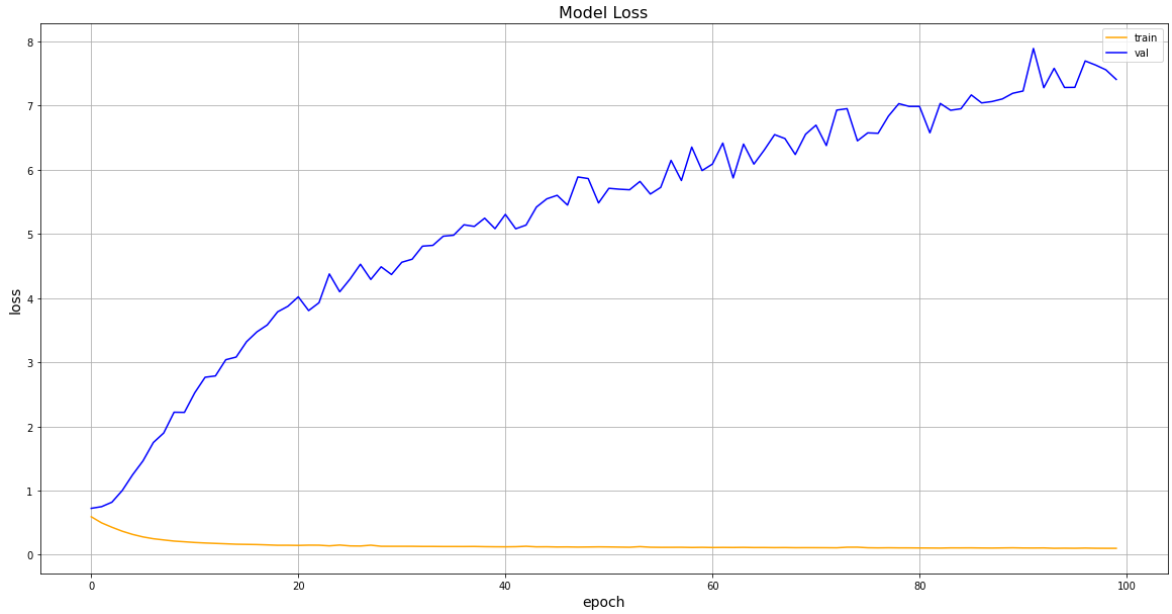
V prvom experimente na základe analýzy k prvotným úpravám pridali ďalšie, naopak niektoré pôvodné boli pozmenené. V prvom experimente bol navyše odstránený odstránený stĺpec žánru ***Free to Play,*** trénovacie dáta sme vybalansovali extrakciou rovnakého počtu platených a bezplatných titulov, oddelili sme stĺpec ***is_free,*** ktorý bude target pri trénovaní, testovaciu množinu sme v rovnomernom pomere bezplatných a platených titulov rozdelili na dve polovice. Následne sme trénovali 5 modelov s aplikovaním rôzneho počtu vrstiev, Dropout-u a Early Stopping-u s nasledujúcimi výsledkami.

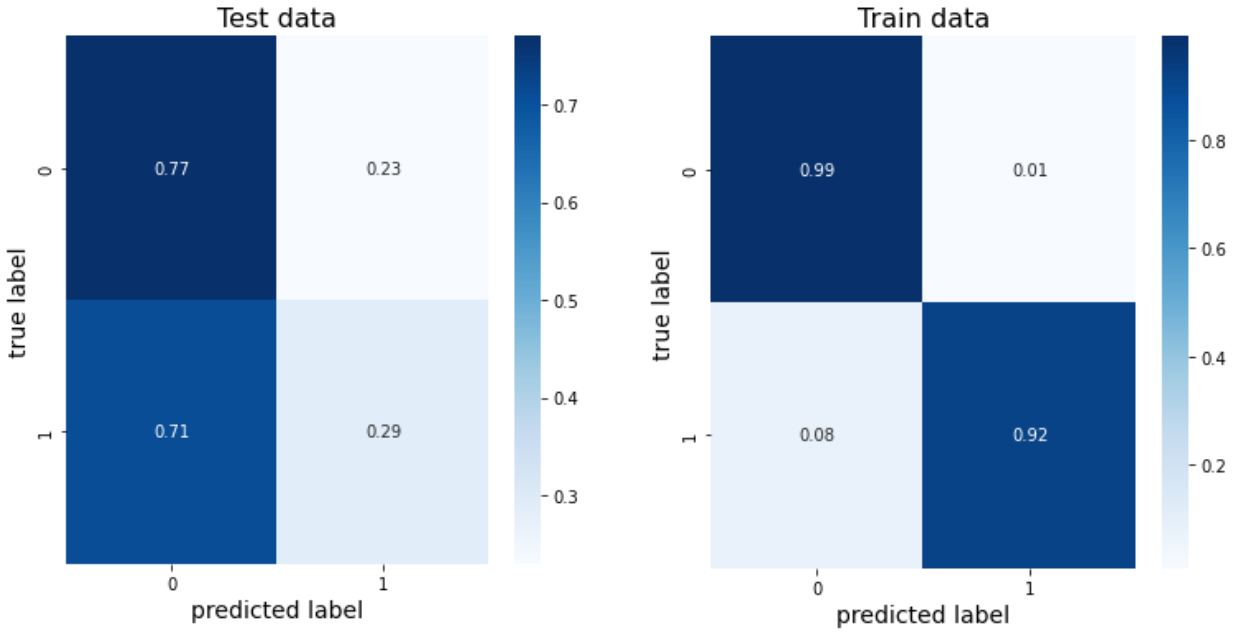
Neurons	Epochs	Activation Function	Loss Function	Optimizer	Dropout	Early Stopping	Train Loss	Train Acc,	Test Loss	Test Acc.
288	100	ReLu	Binary Crossentropy	ADAM – (LR=0.0001)	-	-	0.201	0.935	4.219	0.512
288-144-72	100	ReLu	Binary Crossentropy	ADAM – (LR=0.0001)	-	-	0.100	0.967	2.962	0.530
288-144-72-36-18	100	ReLu	Binary Crossentropy	ADAM – (LR=0.0001)	-	-	0.102	0.966	2.426	0.500
288-144-72-36-18	100	ReLu	Binary Crossentropy	ADAM – (LR=0.0001)	0.5	-	0.129	0.951	2.296	0.516
288-144-72-36-18	1000	ReLu	Binary Crossentropy	ADAM – (LR=0.0001)	0.5	Validation Loss – (50)	0.144	0.956	3.282	0.504

Mikuš Filip

Priebeh trénovania a výsledky najlepšieho (vyznačeného) modelu.







Z výsledkov je možné vidieť, že model trpí pretrénovaním, výsledná schopnosť kategorizovať bezplatné hry v testovacej množine je taktiež problematická.

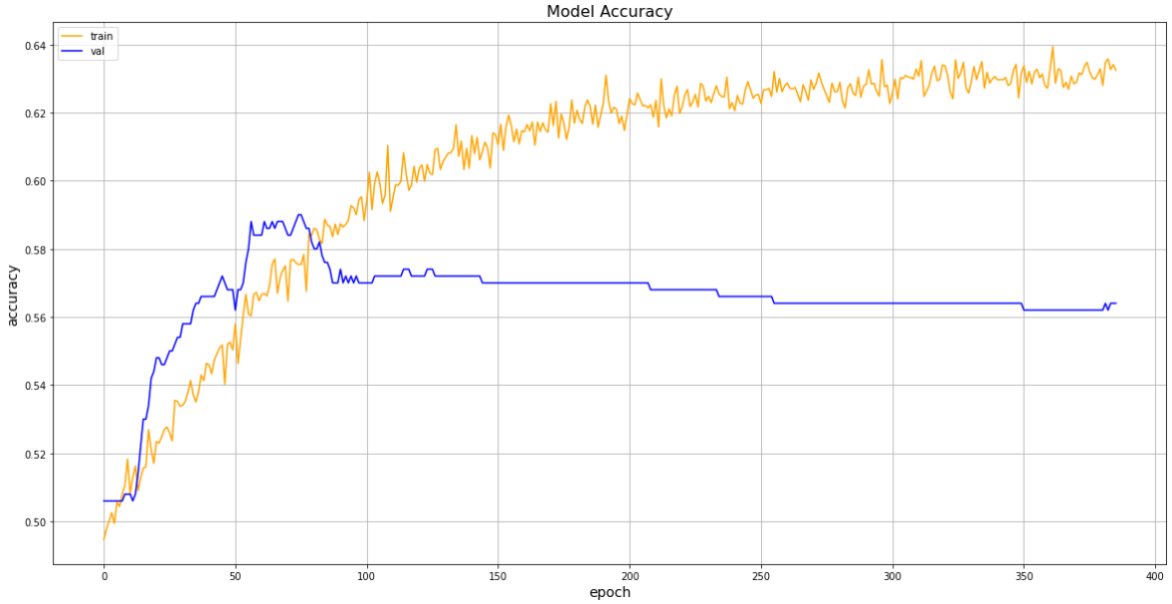
Súbor experimentov č.2

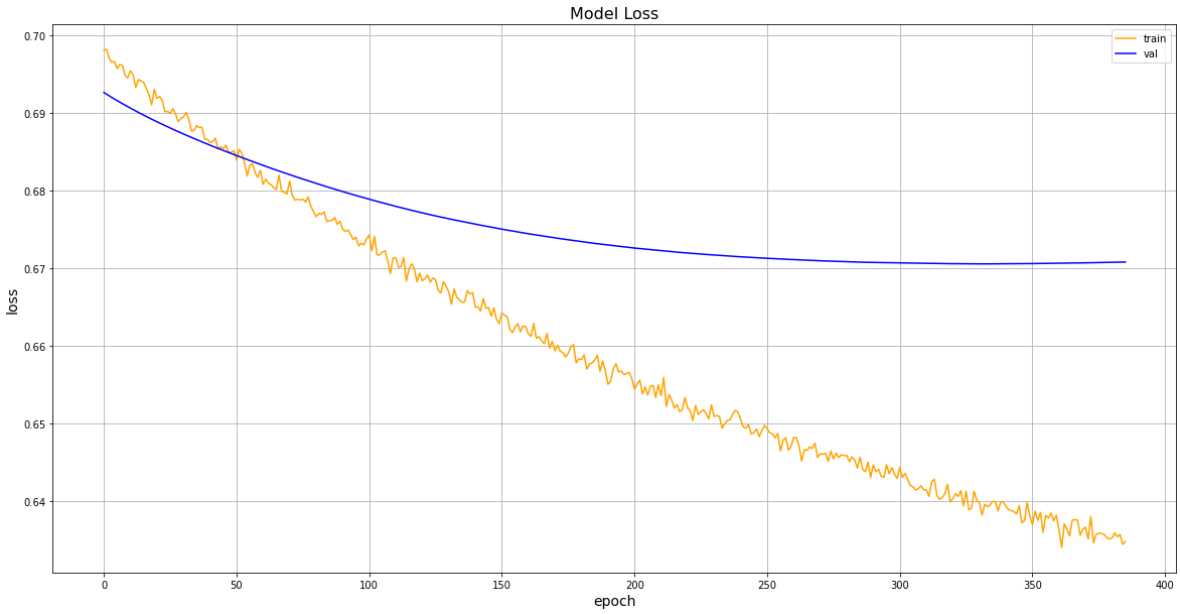
V druhom experimente sme na základe predošlých experimentov a korelačnej matice vyextrahovali stĺpce ***score, is_free, ccu, languages, self_published, is_single_player, is_multi_player, release_year, tags*** a upravili sme learning rate na hodnotu ***0.0000001***. Z tagov bol vyextrahovaný najpočetnejší a bola mu pridelená hodnota z intervalu. Ostatné úpravy boli zachované.

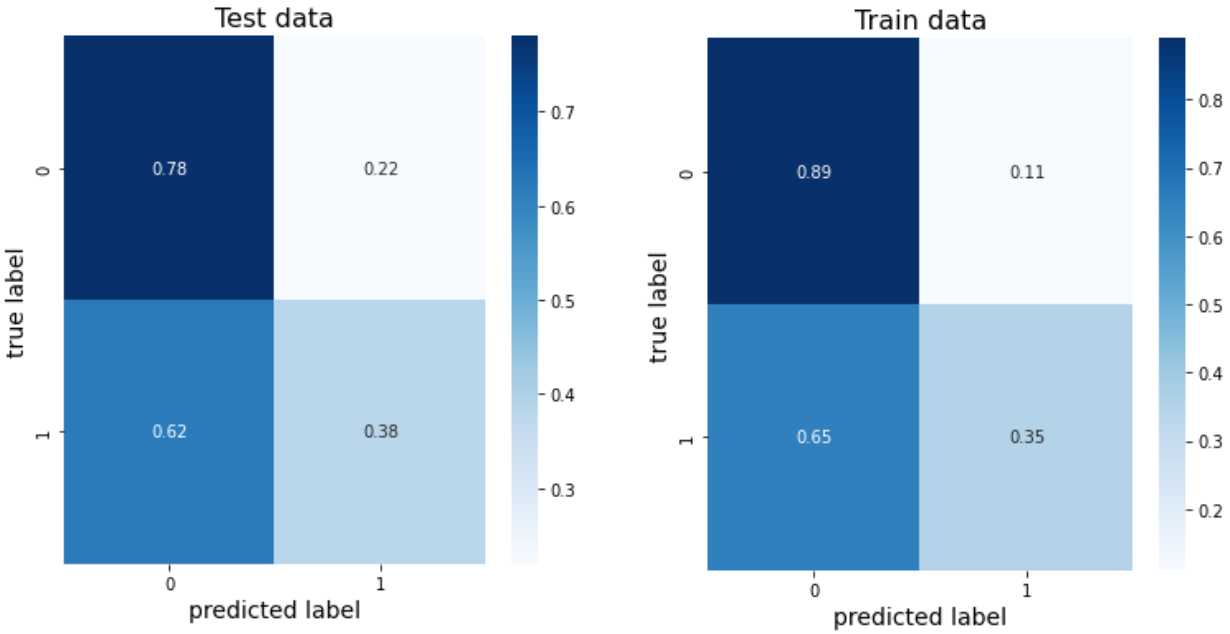
Neurons	Epochs	Activation Function	Loss Function	Optimizer	Dropout	Early Stopping	Train Loss	Train Acc,	Test Loss	Test Acc.
288	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	-	-	0.691	0.517	0.691	0.524
288	100	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	-	0.601	0.679	0.676	0.566
288	1000	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	Validation Loss – (50)	0.635	0.632	0.668	0.578
288-144-72	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	-	-	0.570	0.749	0.687	0.580
288-144-72	100	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	-	0.246	0.926	1.922	0.562
288-144-72	1000	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	Validation Loss – (50)	0.607	0.683	0.674	0.560
288-144-72-36-18	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	-	-	0.292	0.907	1.172	0.580
288-144-72-36-18	100	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	-	0.240	0.930	2.090	0.554
288-144-72-36-18	1000	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	Validation Loss – (50)	0.603	0.707	0.674	0.572

Mikuš Filip

Priebeh trénovania a výsledky najlepšieho (vyznačeného) modelu.







Z výsledkov je zrejmá menšia miera pretrénovania, schopnosti modelu na testovacej množine sú stále nedostatočné, výsledná schopnosť kategorizovať bezplatné hry v testovacej množine je opäť problematická.

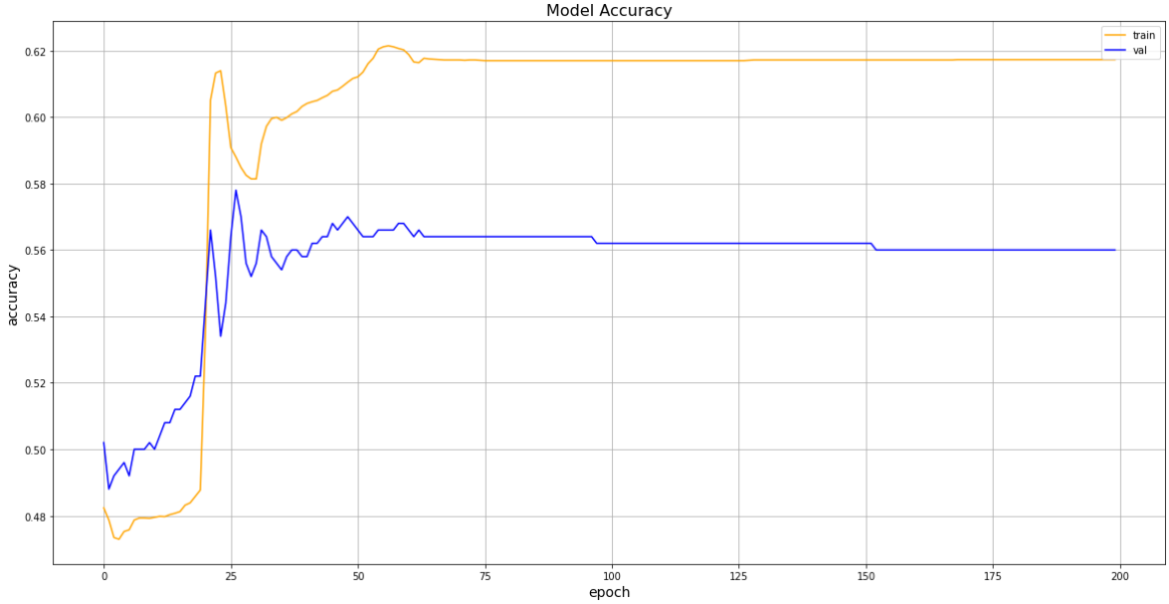
Súbor experimentov č.3

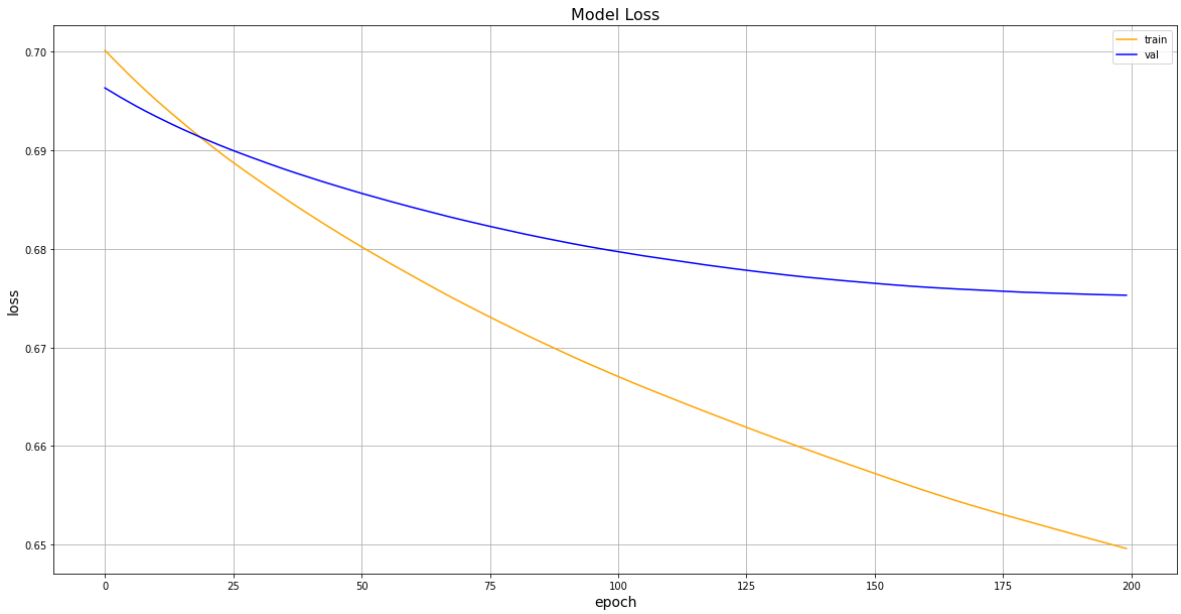
V treťom experimente sme sme na základe predošlých experimentov vyextrahovali stĺpce ***score, is_free, ccu, languages, self_published, is_single_player, is_multi_player, release_year a owners.*** Ostatné zmeny zostali zachované.

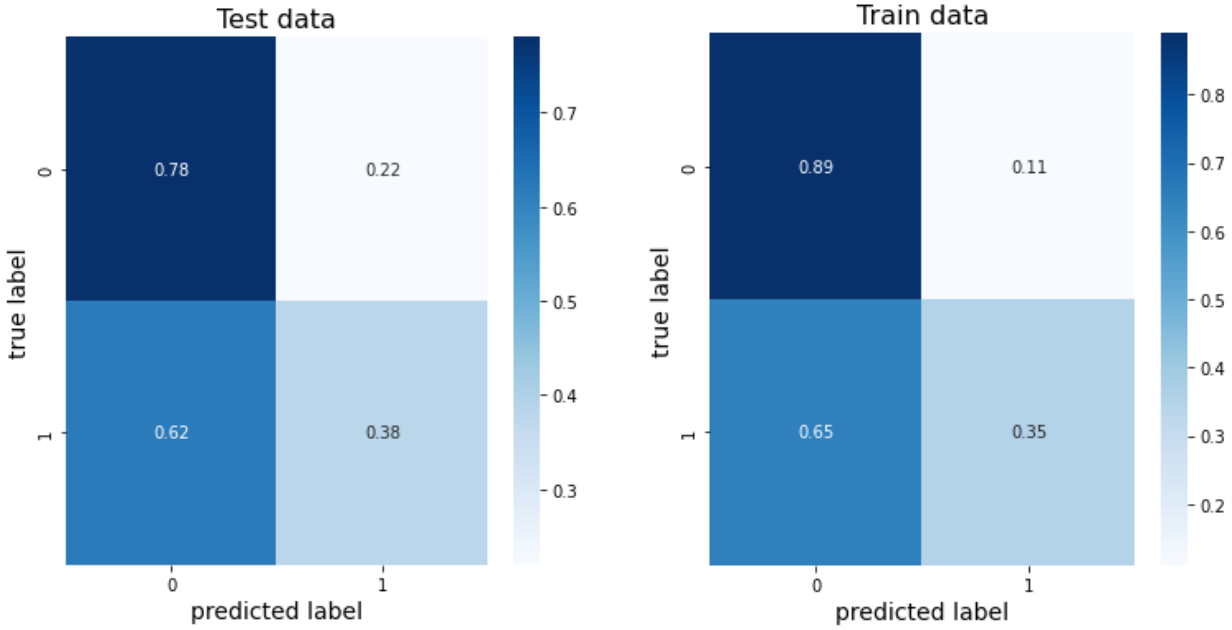
Neurons	Epochs	Activation Function	Loss Function	Optimizer	Dropout	Early Stopping	Train Loss	Train Acc,	Test Loss	Test Acc.
288	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	-	-	0.647	0.617	0.676	0.578
288	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	-	0.662	0.616	0.680	0.558
288	1000	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	Validation Loss – (50)	0.648	0.618	0.679	0.544
288-144-72	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	-	-	0.603	0.632	0.688	0.534
288-144-72	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	-	0.615	0.646	0.681	0.546
288-144-72	1000	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	Validation Loss – (50)	0.633	0.628	0.678	0.570
288-144-72-36-18	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	-	-	0.587	0.691	0.688	0.546
288-144-72-36-18	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	-	0.603	0.699	0.681	0.544
288-144-72-36-18	1000	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	Validation Loss – (50)	0.605	0.702	0.679	0.544

Mikuš Filip

Priebeh trénovania a výsledky najlepšieho (vyznačeného) modelu.







Z výsledkov je zrejmá menšia miera pretrénovania, schopnosti modelu na testovacej množine sú stále nedostatočné, výsledná schopnosť kategorizovať bezplatné hry v testovacej množine je opäť problematická.

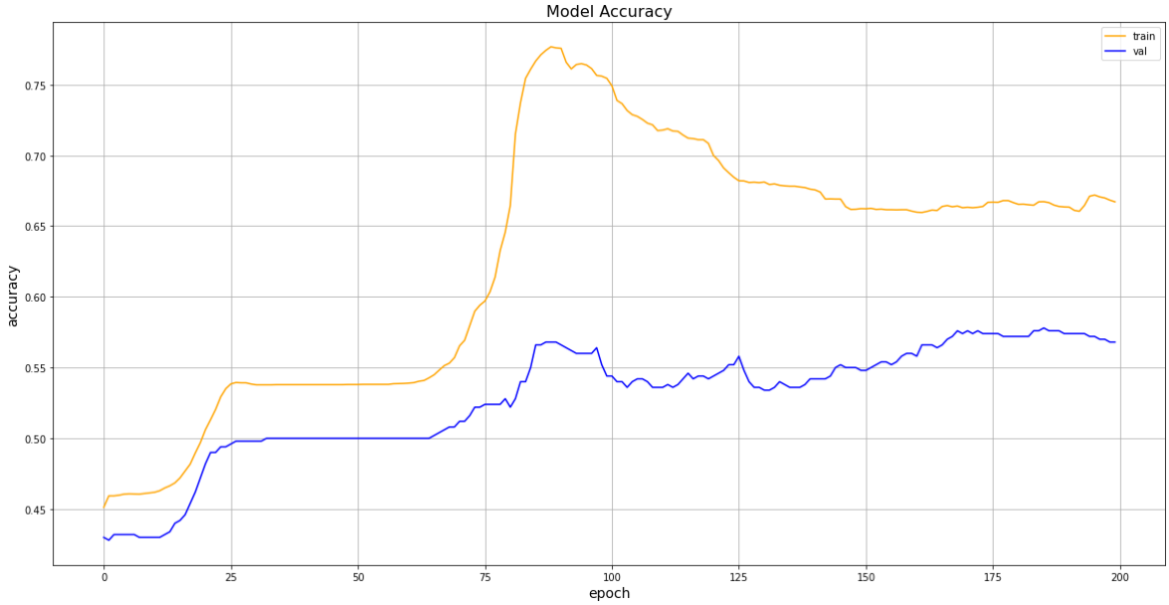
Súbor experimentov č.4

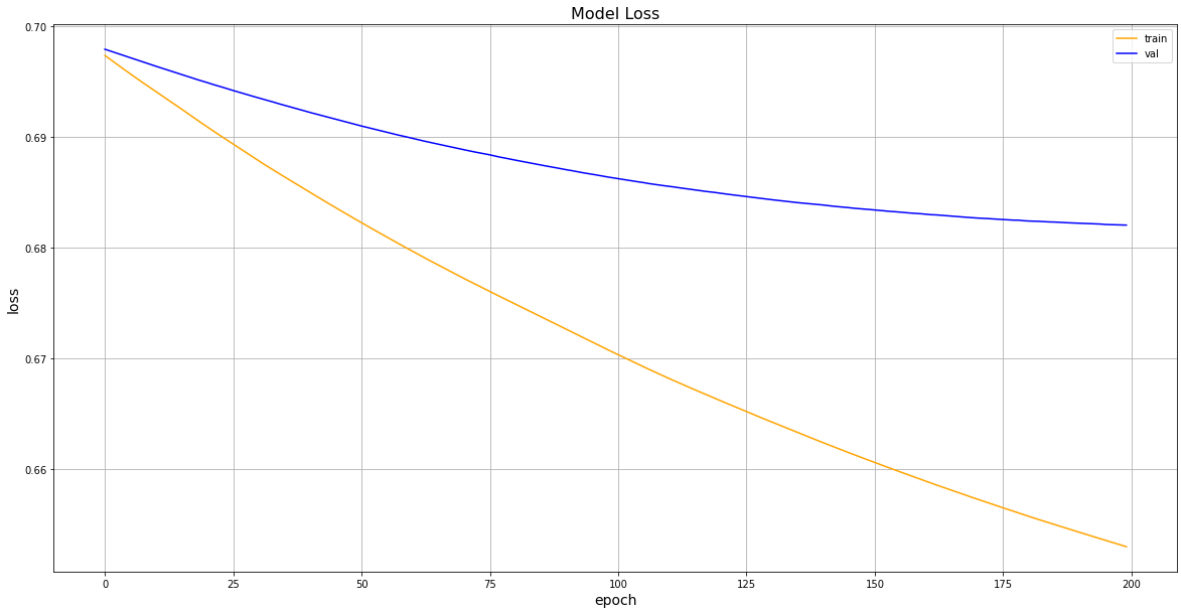
V štvrtom experimente sme sme na základe predošlých experimentov rovnako vyextrahovali stĺpce ***score, is_free, ccu, languages, self_published, is_single_player, is_multi_player, release_year a owners.*** Kvôli zlému pomeru predikcii v konfúznych maticiach sme umelo odobrali 700 vzoriek platených titulov. Ostatné zmeny zostali zachované.

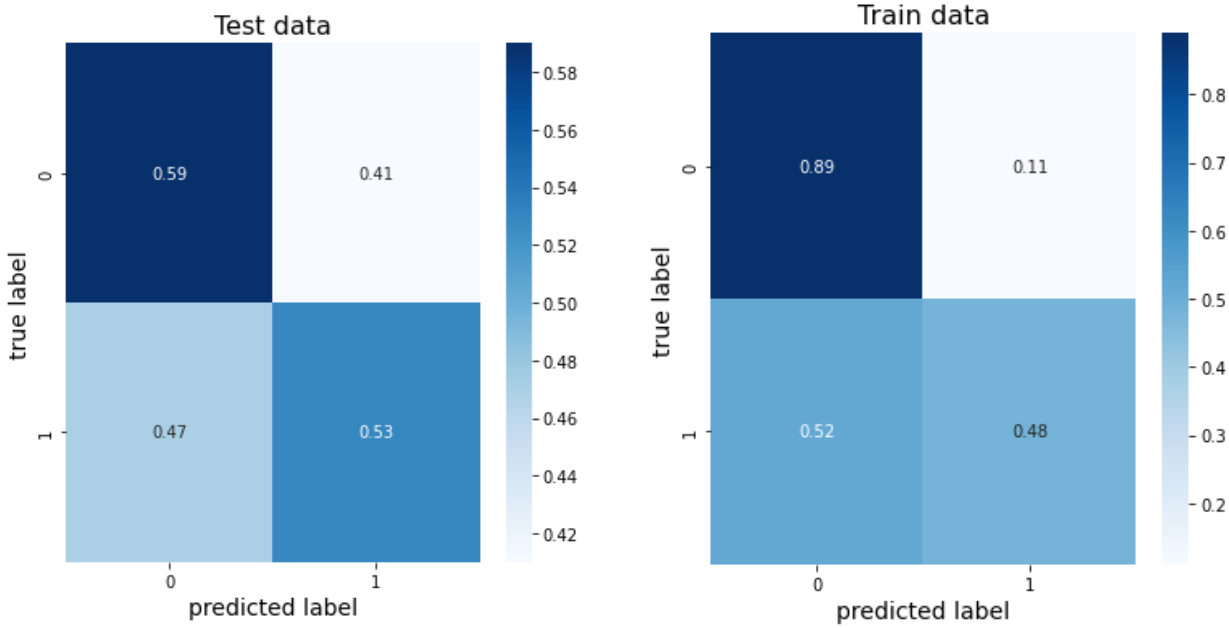
Neurons	Epochs	Activation Function	Loss Function	Optimizer	Dropout	Early Stopping	Train Loss	Train Acc,	Test Loss	Test Acc.
288	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	-	-	0.653	0.667	0.681	0.558
288	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	-	0.665	0.596	0.681	0.548
288	1000	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	Validation Loss – (50)	0.648	0.607	0.679	0.546
288-144-72	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	-	-	0.604	0.690	0.696	0.540
288-144-72	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	-	0.610	0.708	0.701	0.546
288-144-72	1000	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	Validation Loss – (50)	0.633	0.670	0.684	0.552
288-144-72-36-18	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	-	-	0.573	0.767	0.713	0.552
288-144-72-36-18	200	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	-	0.608	0.719	0.682	0.560
288-144-72-36-18	1000	ReLu	Binary Crossentropy	ADAM – (LR=0.0000001)	0.5	Validation Loss – (50)	0.633	0.658	0.680	0.552

Mikuš Filip

Priebeh trénovania a výsledky najlepšieho (vyznačeného) modelu.







Z výsledkov je zrejmá menšia miera pretrénovania, schopnosti modelu na testovacej množine sú stále nedostatočné, výsledná schopnosť kategorizovať bezplatné hry v testovacej množine je však prijateľnejšia.