

Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky



Modely logistické regrese v oblasti esportových dat

BAKALÁŘSKÁ PRÁCE

Studijní program: Aplikovaná informatika

Studijní obor: Aplikovaná informatika

Autor: Michal Lauer

Vedoucí práce: Ing. Zdeněk Šulc, Ph.D.

Praha, Duben 2022

Prohlášení

Prohlašuji, že jsem bakalářskou práci *Modely logistické regrese v oblasti esportových dat* vypracoval samostatně za použití v práci uvedených pramenů a literatury.

V Praze dne DD. Dubna 2021

.....

Podpis studenta

Poděkování

Rád bych poděkoval panu doktoru Zdenku Šulcovi, který mou bakalářskou práci podpořil a vedl, i přes odlišný studijní obor. Dále děkuji autorům knih, jmenovitě ... , za poskytnutou příležitost se ve logistických modelech zlepšit. Bez nich by se práce psala velmi složitě.

Abstrakt

Práce se zabývá predikcí výsledku esportových zápasů dle několika proměnných. V prvním rámci bakalářské práce je v krátkosti představen esport a problematika, kterou se práce zabývá. Dále jsou zde představeny důležité termíny a pojmy, které jsou klíčové ke správné interpretaci dat a výsledků. V druhé části, která se zabývá teorií, jsou představeny metody k popisu dat, modely, a hodnocení kvality modelů. Také zde nalezneme popisy grafů, které jsou použity k shrnutí a popisu datového souboru. V poslední části jsou metody použity v praxi. Nejprve je představen dataset a jeho proměnné. Z proměnných je vybráno pouze několik hlavních prediktorů, které jsou následně použity k výsledné predikci. Dále se zde nachází popis prediktorů a pomocí grafů a slovní interpretace. V závěru je zde sestaven multivariabilní logistický regresní model, který předpovídá výsledek zápasů. Ten je vyhodnocen pomocí již zmíněných statistik. Na konci práce je zhodnocení výsledných predikcí. Nachází se zde i zamyšlení nad použitím modelu v reálném životě a jeho možná vylepšení v budoucnu.

Klíčová slova

Model, logistická regrese, predikce, esport

Abstract

– Bude přeložen po odsouhlasení abstraktu v češtině

Keywords

Model, logistic regression, prediction, esport

Obsah

1	Úvod	3
2	Představení esportu	4
2.1	Historie esportu	4
2.2	Zasazení do dnešní doby	4
2.3	Představení titulu Counter-Strike: Global Offensive	6
2.4	Propojení práce a titulu Counter-Strike: Global Offensive	7
2.4.1	Vysvětlení hry	7
3	Teoretická část	9
3.1	Vizualizace dat	9
3.1.1	Bodový graf	9
3.1.2	Histogram	9
3.1.3	Boxplot	9
3.2	Logistický regresní model	10
3.2.1	Jednoduchý logistický regresní model	10
3.2.2	Vícerozměrný logistický regresní model	10
3.2.3	Optimalizace parametrů	10
3.2.4	Vyhodnocení modelu	10
4	Praktická část	11
4.1	Metodika	11
4.2	Představení data setu	11
4.3	Příprava data setu	11
4.4	Omezení dat	11
4.5	Sestavení modelu	11
4.5.1	Interpretace	11
4.6	Vyhodnocení modelu	11
4.6.1	Statistika 1	11
5	Závěr	12
	Seznam použité literatury	13
	Seznam elektronických zdrojů	14
	Seznam obrázků	15
	Seznam tabulek	16
	Seznam použitých zkratk	17

1. Úvod

Esport je jedno z nejrychleji rostoucích odvětví v dnešní době. V roce 2021 se jeho tržní hodnota pohybovala kolem jedné miliardy dolarů - skoro 50% nárůst oproti roku 2020. Dle portálu statista.com lze předpovídat, že v roce 2024 esport překročí hodnotu 1,5 miliardy dolarů (1). Dalo by se spekulovat, že za takový velký nárůst je zodpovědná aktuální pandemie. Většina populace, hlavně ta mladší, je nucena zůstat doma. Toto otevřelo dveře se s esportem přirozeně seznámit a nějakým způsobem se ho účastnit (*online divák, soutěžící, organizátor, fanoušek...*). Esport je vlastně sport, akorát s počítačovými hrami. Hrají se různé kategorie her - střílečky (*Counter-Strike: Global Offensive (CS:GO), Valorant*), arény (*League of Legends (LoL)*), či karetní hry (*Hearthstone (HS)*).

Toto téma jsem si zvolil hlavně kvůli tomu, že se o oblast zajímám od mého mládí. Když jsem si vybíral téma na bakalářskou práci, chtěl jsem propojit statistiku s něčím, co mě baví a naplňuje - toto je ideální kombinace. Zároveň bylo mým cílem vytvořit práci, která bude v dnešní době relevantní. Zvolené téma je dle mého názoru velmi aktuální, avšak ne pro širokou veřejnost, nýbrž pouze pro lidi, co se o zajímají o esport či sázení. Podobné logistické modely, avšak velmi složitější, mohou totiž sloužit například k vyhodnocení sázkových kurzů. Esport, tak jak klasický sport, je se sázením propojen.

Finální cíl práce je vytvořit logistický model, který předpovídá výsledek zápasů. Tento model je vyhodnocen různými klasifikacemi pro vyhodnocení kvality. Práce také popisuje grafické metody vizualizace dat a teorii k tvorbě a vyhodnocení logistických modelů. Také zde najdeme popis datového souboru a postup, jakým byli vybráni nejvýznamnější prediktory - ať už statisticky, či čistě ze znalosti esportu.

2. Představení esportu

2.1 Historie esportu

I přes fakt, že esport není obecně známý pojem mezi širokou veřejností, má přes 70 let bohaté historie. Za jeho počátky by se daly považovat arkádové automaty, kde hráči z počátku soutěžili sami proti sobě. Největší rozvoj arkádových automatů se děl kolem 70 let minulého století. Nejen za tímto účelem byla 9. 2. 1982 založena *Twin Galaxies National Scoreboard (TGNS)*. TGNS měla na starosti nejen udržování výsledkové tabulky (*ang. scoreboard*), ale i tvorbu prvotních pravidel pro férovou hru. Za tímto účelem byla vydána kniha *Twin Galaxies' Official Video Game & Pinball Book of World Records*.

Na přelomu osmdesátých let minulého století se začal esport vyvíjet již více profesionálním směrem. V roce 1972 pořádala Stanfordská Universita historicky první esportový turnaj v arkádové hře *Spacewar!*. Výherce si mohl odnést předplatné magazínu Rolling Stones. Dále v roce 1983 byl založen první esportový profesionální team, který se nacházel ve Spojených státech. Všechno toto se stalo díky podnikateli Walteru Day, který je jak zakladatel společnosti TGNS a založil již zmíněný první esportový team. Ač se Walter považuje za jednoho z hlavních pionýrů esportu, v roce 2010 TGNS opustil kvůli své vášni pro hudbu.

Další důležitou kapitolou ve vývoji esportu je příchod internetu a výkonných počítačů. Hráči měli rychlejší sestavy, stolní počítače byli cenově dostupnější a díky tomu se dostali k více lidem. Klesala cena hardwaru, vývoj nové technologie a her se zrychloval. Díky rozvoji počítačových sítí se mohli hrát LAN party¹ či organizovat BYOC turnaje². Dále už esport potřeboval jen čas na organický růst a dnes má tržní hodnotu přes jednu miliardu amerických dolarů (1), (2).

2.2 Zasazení do dnešní doby

Jak již bylo zmíněno, esport je v dnešní době téměř miliardová záležitost. Díky pandemii, která trvá již třetím rokem, si esport ještě přilepšil. Dle průzkumu (3) z října roku 2020 si 73 % dotázaných myslelo, že se úroveň zájmu (*ang. level of investment*) a obchodní činnost (*ang. deal activity*) v Q4 2020 a Q1 2021 zvětší. Respondenti, kteří se průzkumu zúčastnili, jsou považováni za „industry professionals“. Tento průzkum byl následně podpořen růstem že tržní hodnoty esportu a mezi lety 2019 a 2020 vzrostla o téměř 50 % (1).

¹Hráči hrají v jedné místnosti na lokální počítačové síti.

²z ang. Bring Your Own Computer, kde si hráči si na akci donesou vlastní počítač

K takto prudkému růstu tržní hodnoty esportu z velké části přispěla právě pandemie. Mladá generace byla nucena zůstat doma, což dovolilo i esportem nedotčeným jedincům do tohoto světa proniknout. Větší zájem o esport přinesl i větší tržby herním studiím, které začali do esportových turnajů investovat více peněz(4)(5). S větším počtem diváku roste i marketingový potenciál, investiční příležitost a kariérní růst.

V dnešní době má esport mnoho titulů, proto představím jen ty nejvýznamnější. Největší esport rivalita je mezi herním titulem LoL a Dota 2. Oba tituly jsou žánru Multiplayer online battle arena (MOBA), díky čemuž mají podobnou, avšak velmi rozdílnou fanouškovskou základnu. Historie mezi tituly je velmi složitá, avšak mimo rozsah této práce. Pro rozšíření znalosti mohu doporučit videa z youtubového kanálu theScore esport o tomto tématu - [The Story of Dota 2](#) a [The Story of League of Legends](#).

Druhý dominantní žánr je First-person shooter (*střílečka z pohledu první osoby*) (FPS). V této kategorii dominují hry CS:GO a Valorant. Zde proti sobě hrají dva týmy, většinou složené z pěti hráčů. Každý hráč pak má v týmu různou roli, jako např. velitel či odstřelovač. Jeden tým má obvykle za úkol něco zničit (*položit bombu, unést rukojmí*) a druhý tým jim v tom musí zabránit (*ochránit oblast proti bombě, záchrana rukojmí*).

Poslední žánr, který zmíním, je Battle Royale (*hra o přežití*) (BR). V těchto hrách hraje buď každý hráč za sebe, ve dvojicích, nebo skupinách po čtyřech. Zde hráči padají na začátku kola na velkou mapu. Jejich úkolem je získat tzn. „loot“ (*vybavení*), aby mohl porazit ostatní hráče a kolo sami, nebo s týmem vyhrát. Nacházejí se zde různé role, avšak trochu rozdílné oproti žánru FPS. Hlavním titulem této kategorie je hra Fortnite, která žánru dominuje. Stal se z ní jak esport titul, tak perfektní marketingové místo pro teenagery. Hráči si zde mohou koupit oblečky různých filmových či komiksových postav. Pokud vychází nový film, ve hře se může objevit „event“ (*událost*), který daný film propaguje. Toto lze vidět například na propagaci [Avengers: Endgame](#).

2.3 Představení titulu Counter-Strike: Global Offensive

Counter-Strike: Global Offensive jak ho známe dnes, má bohatou a dlouhou historii. Ne vždy se to ovšem jmenovalo stejně. Úplně první iterace hry se jmenovala čistě Counter-Strike a byl to pouze mód³ do hry Half-Life. Half-Life byl vyvinutí společností Valve, tehdy primárně společností zaměřenou na vyvíjení her. Mód byl vytvořen studenty vysoké školy, panem *Minh Le* a *Jess Cliffe*. Toto rozšíření začali programovat v roce 1999. Jelikož mód byl neoficiálním rozšířením, Valve se o něj moc nezajímalo. Až po pěti betaverzích hry Counter-Strike si společnost Valve všimla rozšíření, její komunity, ale především jejich autorů. Minh a Jess se v roce 2000 stali oficiálními zaměstnanci Valve, prodali „duševní vlastnictví“ módu Valve. Autoři, nově jako zaměstnanci Valve, roku 2000 vydávají první oficiální verzi hry Counter-Strike. I přes toto „oficiální“ datum vydání je většina komunity přesvědčena, že výročí má CS:GO v den svého úplně první vydání, a to 18. června 1999.

Hra je z žánru First-person shooter (*střílečka z pohledu první osoby*) a hraje se primárně online proti skutečným hráčům. Counter-Strike se v herní komunitě rychle rozrostl jeho jednoduchostí. Hra se dá velmi dobře popsat pořekadlem „*Lehké hrát, těžké vypilovat*“ (*ang. Easy to play, hard to master.*). Hra má mechaniky⁴, které jsou lehké na pochopení, ale velmi těžké na vypilování k dokonalosti. Spolu s touto vlastností je hra vlastně velmi jednoduchá a hráč hraje buď za policisty, nebo za teroristy. Hráči tak mohli, a stále můžou, hru velmi lehce a rychle začít hrát - tento formát se totiž za posledních 20 let nezměnil.

Hra tedy rostla zejména díky své komunitě. Hráči hru různě upravovali, přidávali další módy, typy her, zbraně, mapy či audiovizuální obsah. Tento trend se přenášel přes mnoho různých verzí hry. První velký „průlom“ udělala verze 1.6, tedy Counter-Strike 1.6. Ta kvetla jak esportem, tak komunitním obsahem. Jen v České a Slovenské republice bylo několik herních serverů, na kterých se mohlo sejít sta tisíce hráčů. Např. na česko-slovenském herním portálu *kotelna* hrálo celkem přes 1,5 milionu unikátních hráčů (6). Hra byla populární nejen mezi „casual“ hráči, ale i profesionály.

Counter-Strike 1.6 je pionýrem esportu pro FPS žánr. Za podpory Valve se hráli první major⁵ turnaje, kde hráči mohli ukázat svůj um za tehdy relativně velkou sumu peněz. Hra se časem vyvíjela, hráči nalézali nové strategie či triky a Valve vydalo novou verzi - Counter-Strike: Source. Tato nová verze získala nepříliš pozitivní ohlas, jelikož velmi rozdělila herní komunitu. Představila nové mechaniky, staré mechaniky změnila a hráčům, zejména v esportu, se nechtělo učit něco úplně nového. Valve se rozhodlo sjednotit herní komunitu, a proto vydalo novou verzi hry - Counter-Strike: Global Offensive

³upravení či rozšíření hry

⁴herní prvky či unikátní vlastnosti

⁵turnaj pořádaný přímo Valve, který má největší prestiž

CS:GO se snažilo sjednotit oba tábory - Counter-Strike 1.6 a Counter-Strike: Source. Hra vyšla 21. srpna 2012 a z počátku nebyla tolik úspěšná, ale díky přidání různých skinů(7) na zbraně hra přilákala úplně nové publikum. Díky novému a velkému publiku se začali hrát menší esportové turnaje právě ve hře CS:GO, ke kterým se později přidali i profesionále z předchozích dvou verzí. Díky tomuto organickému růstu má Counter-Strike velmi silnou komunitu, která se o hru i nadále stará. I přes netradiční interakci mezi Valve a herní komunitou hra stále roste. CS:GO se díky své dlouhé historii, bohaté komunitě a různým možnostem, jak hru hrát, dostala na špičku esportu. I přes několik titulů, které se s hrou snaží soutěžit, je hra stále největším a nejsledovanějším esport titulem v rámci FPS žánru(8).

2.4 Propojení práce a titulu Counter-Strike: Global Offensive

2.4.1 Vysvětlení hry

Jak již bylo zmíněno, CS:GO hraje pět hráčů proti pěti (*dále jen 5v5*). Hra se většinou hraje online, avšak velké esportové turnaje se hrají offline, tedy v nějaké např. aréně. Hra má v základu 30 kol a po prvních patnácti se mění strany. Jedna strana jsou policisté (*Counter-Terrorists* či *CT*), kteří mají za úkol chránit „bomboviště“ - část mapy, která má vybuchnout. Naopak cíl Teroristů (*T*) je právě bombu položit a „bomboviště“ nechat vybuchnout. Vyhrává team, který první dosáhne 16 kol. Pokud ovšem po prvních 30 kolech je stav nerozhodný, tedy 15:15, hraje se prodloužení. Tento formát není standardizovaný pro všechny turnaje, proto se budu zaměřovat čistě na turnaje, které pořádá Valve (*již zmíněné a nejvíc prestižní Majory*). Zde se hraje prodloužení ve formát Bo6, tedy kdo první získá 4 body, vyhraje zápas. Takto může jít zápas teoreticky do nekonečna. Nejdelší semi-profesionální zápas, který se ovšem neodehrál na Majoru, se stal mezi týmem exceL a XENEX(9). Zápas pokračoval do úctyhodných 88 kol.

V každém kole má tým určitý počet peněz. Každá hráč začíná polovinu (*ted v první a šestnácté kolo*) s \$800. Finance každého hráče pak záleží na mnoha faktorech - kolik vyhrál jeho team kol v řadě, kolik nakoupil zbraní, kolik zabil nepřítelů, kolik peněz dostane hráč za zabití či jak kolo skončí. V profesionálním teamu je velmi obtížné pracovat s financemi, jelikož hráči musí být s financemi na jedné stránce. V tuto chvíli přichází na řadu tzn. In-Game Leader (*velitel teamu*). Tuto roli má většinou jeden hráč v každém teamu. Je to ta nejdůležitější role ze všech - má na starosti finance, rozhoduje kdy se koupí a kdy půjde tzn. eco (*hráči nekoupí nic, aby ušetřili peníze*), jaké se budou hrát mapy či jaká se půjde v daném kole strategie. V dnešní době k tomu In-Game Leader má i pomocníka - trenéra. Ten hru nehraje, ale pozoruje hráče a dává jim různé typy a triky.

Role trenéra není nijak silně definovaná a každý esportový team má trochu jiného trenéra. V jednom případě může být trenér čistě jako podpora - pomáhá hráčům když se nedaří a řeší interní problémy. V jiném teamu může ovšem mít velký zásah do hry, pomáhat In-Game Leaderovi se strategiemi, obelstění soupeře či sledováním předchozích zápasů pro kontinuální zlepšování teamu. Další role v teamu jsou například Entry Fragger (*má za úkol získat první zabití pro team*), support (*podporuje svůj team s pomocí různých granátů nebo se často pro svůj team obětuje*), AWP hráč (*hráč je specifický tím, že hraje primárně s jednou zbraní - odstřelovací puškou AWP*) a Lurker (*chodí po mapě sám a snaží se nepřítele odchytnout ze stran, které by nečekali*)

Zápasy se pak hrají ve formátech „Best of“. *Best of 3* například znamená, že se hrají tři mapy - kdo vyhraje dvě, vyhrál zápas. Turnaje se pak odehrávají v tradičních formátech, jako je pavouk. Ten se charakterizuje tím, že vypadá jak pavučina, jde zleva a každý team může prohrát pouze jednou. Následně tu máme Upper/Lower bracket formát, který je v podstatě „pavoučí formát“, akorát jsou zde dvě sítě a každý team může prohrát maximálně jednou, jinak je vyřazen. Specifičtější formát pro CS:GO je například swiss, který je složitější a mimo rozsah této práce.

Práce se zabývá právě predikcí daného zápasu. K tomuto lze použít např. hodnocení hráčů či umístění teamu na světovém žebříčku. Dále také záleží, na jaké mapě se zápas hraje. Když se budeme dívat na historické zápasy, musíme se dívat pouze na to, kdo zápas vyhrál - ne na rozdíl kol. Ten může být velmi zavádějící a zápas s výsledkem 16:7 mohl být více vyrovnaný než zápas s výsledkem 16:13. Toto jde vidět na zkušenostech teamu. Zápas mezi velmi dobrými esportovými teamy může dopadnout 16:7, ale může být velmi těsný - oba dva týmy hráli dobře a o výsledku rozhodovali maličkosti. Pokud hrají dva méně zkušené týmy, zápas může skončit např. 16:13, ale nemusí být vůbec blízko - jeden team může dělat zbytečné chyby (*které by lepší team neudělal*), které se normálně nedějí, což může vést k více těsnému výsledku i přes rozdílné hodnocení teamů.

3. Teoretická část

V následující části jsou popsány jak teoretické metody pro vizualizaci dat, tak i tvar, forma a vyhodnocení logistického regresního modelu. Ke každé části, která se věnuje popisu dat pomocí nějakého grafu, je přidána praktická ukázka s popisem a praktickým vysvětlením. U regresního modelu je následně dán důraz na teorii, jelikož model je následně prakticky použit v celé třetí části. Důležité je si uvědomit, že nástrojů pro vizualizaci dat je mnoho a grafů či souhrnných statistik je nespočet. V této práci se ovšem věnujeme vizualizacím, které jsou pro regresní model vhodné. Testovací citace: (10), (11)

3.1 Vizualizace dat

... Dopsat

3.1.1 Bodový graf

... Dopsat

3.1.2 Histogram

... Dopsat

3.1.3 Boxplot

... Dopsat

3.2 Logistický regresní model

... Dopsat

3.2.1 Jednoduchý logistický regresní model

... Dopsat

3.2.2 Vícerozměrný logistický regresní model

... Dopsat

3.2.3 Optimalizace parametrů

... Dopsat

Metoda nejmenších čtverců

... Dopsat

Metoda největší pravděpodobnosti

... Dopsat

3.2.4 Vyhodnocení modelu

... Dopsat

4. Praktická část

4.1 Metodika

... Proč jsem si vybral právě vybrané metody a statistiky

4.2 Představení data setu

... Popis dat a struktury

4.3 Příprava data setu

... Popis PROČ a jak byl data set transformován, žádný kód, pouze slovně

4.4 Omezení dat

... Restrikce nad data, výběr jednoho konkrétního teamu

4.5 Sestavení modelu

... Slovní popis modelu a jeho proměnných, rovnice

4.5.1 Interpretace

... Matematická interpretace modelu a výstupu z Rka (koeficienty, odchylky, významnost...)

4.6 Vyhodnocení modelu

4.6.1 Statistika 1

... Různé statistiky, které musím vybrat aa

5. Závěr

Seznam použité literatury

10. HEBÁK, Petr. *Statistické myšlení a nástroje analýzy dat*. 2. vyd. Informatorium, 2015.
ISBN 978-80-7333-118-4.
11. KLEINBAUM, David. *Logistic regression : a self-learning text*. 3. vyd. Springer, 2010.
ISBN 978-1-4939-3697-7.

Seznam elektronických zdrojů

1. GOUGH, Christina. esports market revenue worldwide from 2019 to 2024. *statista*. 2021. Dostupné také z: <https://www.statista.com/statistics/490522/global-esports-market-revenue/>.
2. LARCH, Florian. the history of the origin of esports. *ispo*. 2019. Dostupné také z: <https://www.ispo.com/en/markets/history-origin-esports>.
3. GOUGH, Christina. impact of the coronavirus (covid-19) pandemic on the level of investment and deal activity in esports according to industry professionals worldwide as of october 2020. *statista*. 2021. Dostupné také z: <https://www.statista.com/statistics/1247902/covid-impact-esports-investments/>.
4. PROFESSEUR. esea increase prize pool and number of seasons for 2021; simplify path to pro league. *hltv*. 2021. Dostupné také z: <https://www.hltv.org/news/30926/esea-increase-prize-pool-and-number-of-seasons-for-2021-simplify-path-to-pro-league>.
5. LIQUIPEDIA. pgl major stockholm 2021. *liquipedia*. 2021. Dostupné také z: <https://liquipedia.net/counterstrike/PGL/2021/Stockholm>.
6. CSKO. základní statistiky. *csko.cz*. 2021. Dostupné také z: <https://stats.csko.cz/statsx/hlstats.php>.
7. VALVE. counter-strike: the arms deal update. 2013. Dostupné také z: <http://counter-strike.net/armsdeal>.
8. HENNINGSON, Joakim. the history of counter-strike. *redbull*. 2020. Dostupné také z: <https://www.redbull.com/se-en/history-of-counterstrike>.
9. HLTV.ORG. xenex vs. excel at esl uk premiership season 1. *hltv.org*. 2015. Dostupné také z: https://www.hltv.org/matches/2295340/xenex-vs-excel-esl-uk-premiership-season-1?__cf_chl_jschl_tk__=82S_Uc_dNU8PM71eUY1VKzNUZkp5_ArTb69qteh2wBI-1641459294-0-gaNycGzNChE.

Seznam obrázků

Seznam tabulek

Seznam použitých zkratek

CS:GO Counter-Strike: Global Offensive

LoL League of Legends

HS Hearthstone

BR Battle Royale (*hra o přežití*)

MOBA Multiplayer online battle arena

FPS First-person shooter (*střelečka z pohledu první osoby*)

TGNS Twin Galaxies National Scoreboard

Část I

Přílohy