|  |  |
| --- | --- |
|  | Zbornik radova Fakulteta tehničkih nauka, Novi Sad |

**UDK: 004.85**

**DOI:** [**https://doi.org/10.24867/20BE19Petrovic**](https://doi.org/10.24867/20BE19Petrovic)

**ПРЕДИКЦИЈА ИСХОДА ПРОФЕСИОНАЛНИХ ТИМСКИХ МЕЧЕВА У ИГРИЦИ CS:GO**

**PREDICTION OF THE OUTCOME OF PROFESSIONAL TEAM MATCHES IN GAME CS:GO**

Душан Петровић, *Факултет техничких наука, Нови Сад*

**Област – ЕЛЕКТРОТЕХНИКА И РАЧУНАРСТВО**

**Кратак садржај –** *С обзиром да награде на CS:GO турнирима достижу и по неколико милиона долара, тимовима би значила свака информација која их по­тенцијално може довести до победе на мечу и турни­ру. Систем који би професионалним тимовима указао на недостатке, односно дао предикцију исхода меча узимајући у обзир састав тима и компатибилност њихових стилова игре, био би од велике користи при креирању тимова. Идеја овог рада јесте да кроз неко­лико експеримената покаже како одређени фактори ути­чу на предикцију исхода меча, односно која њи­хова комбинација би дала најбоље резултате предик­ције. Поред успеха професионалних играча током каријере, у обзир би се узимао и фактор државе из које долази играч као и његов стил игре. У раду је такође представљен и алгоритам за процену rating-а играча. За претпоставку стила игре корис­тио се KMeans алго­ритам за кластеровање, док су се за проблем предик­ције исхода меча користили XGBoost Regressor, Ran­dom Forest алгоритми, као и неуронска мрежа. Експе­рименти су у већини случајева давали очекиване ре­зултате с обзиром на природу проблема. У раду је по­казано да подаци о стилу игре и фактору државе иг­рача у већини случајева позитивно утичу на предик­цију исхода меча.*

**Кључне речи:** *CS:GO, ESport, предикција меча, кластеровање, регресија*

**Abstract** – *Considering that the prizes in CS:GO tournaments reach several million dollars each, any information that can potentially lead them to victory in the match and the tournament would mean a lot to the teams. A system that would indicate the shortcomings of professional teams, and give a prediction of the outcome of the match, taking into account the composition of the team and the compatibility of their playing styles, would be of great use when creating teams. The idea of ​​this paper is to show through several experiments how certain factors affect the prediction of match's outcome and which combination of them would give the best prediction results. In addition to the success of professional players during their career, the factor of the country the player comes from as well as his style of play was also taken into account. The paper also presents an algorithm for evaluating player ratings.*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**НАПОМЕНА:**

**Овај рад проистекао је из мастер рада чији ментор је била др Јелена Сливка, доцент.**

*The KMeans clustering algorithm was used to predict the style of play, while the XGBoost Regressor, Random Forest algorithms, and a neural network were used to predict the match's outcome. In most cases, the experi­ments produced the expected results, given the nature of the problem. In the paper, it was shown that data on the style of play and player's country data, in most cases, have a positive effect on the prediction of the outcome of the match.*

**Keywords** *CS:GO, ESport, match prediction, clustering, regression*

**1. УВОД**

Укупни буџет за новчане награде свих *ESports* такмичења у CS:GO игри у 2021. години је износио 20 милиона долара, што је позиционира на друго место у овој категорији. Игра у свом жанру није никад имала конкуренцију због чега се очекује да ће одржати своју популарност.

У традиционалним спортовима, анализа података је значајно променила многе аспекте данашњег профе­сионалног спорта. Иако анализа спортских података позитивно утиче на одабир играча, креирање тимова и формирање стратегије, спортски аналитичари су и даље ускраћени за добар део података, с обзиром на потешкоће за њихово прикупљање из стварног света. Насупрот томе, видео игре су направљене од пода­така, што олакшава саму анализу. Модерне видео иг­ре, укључујући и CS:GO, прикупљају огромне коли­чине података током мечева, што представља одличну полазну тачку за даљу анализу и истраживање.

С обзиром да је и у *ESport*-овима присутан људски фактор, анализом података из мечева и њиховом обрадом и те како се може доћи до закључака који ће позитивно утицати на формирање тима и повећање шанси за победу. Систем који би предвидео исход меча за различите услове би потпомогао доношење одлука за професионалне играче и повећао њихову шансу за победу.

У овом раду ће бити представљено решење за предикцију исхода професионалног CS:GO меча. Кроз неколико експеримената покушано је да се прикаже како одређени фактори утичу на предикцију исхода меча. Решење ће бити реализовано коришћењем неколико скупова података које садрже информације о мечевима и статистикама играча прикупљених током њихове каријере.

Такође, биће коришћени различити алгоритми за предикцију, где ће се изабрати најбољи од њих након њиховог поређења. Кроз један од експеримената биће дат и предлог решења за процену *rating*-а играча, а затим и поређење са оценом играча из коришћеног скупа података.

За одређивање стила игре употребљен је *KMeans* алгоритам за кластеровање на основу играчевих статистичких података током каријере. За проблем предикције исхода меча коришћени су *XGBoost Regressor* и *Random Forest* алгоритми, као и неуронска мрежа. Најбоље резултате од 62.79% тачности је постигао *XGBoost Regressor.*  Експерименти су у већини случајева давали очекиване резултате, с обзиром на природу проблема, јер се ради о предикцији где је присутан људски фактор и некад су нијансе те које одлуче победника меча, и неретко се дешава да објективно лошији тим победи онај бољи.

У наредном поглављу детаљније ће бити представљена постојећа литература. Затим ће у поглављу 3 бити описана методологија и спроведени екперименти. Након тога, у поглављу 4 се налази дискусија о резулттима. На крају, у поглављу 5, дата је сумаризација рада са предлозима за даље унапређење решења.

**2. ПРЕТХОДНА РЕШЕЊА**

Рад [1] анализира модел који урачунава само тимске одлуке. Циљ рада је реализација модела за предикцију вероватноће победе меча на мапи, на почетку сваке рунде, урачунавајући тимске одлуке у току меча. У вероватноћу победе урачунат је тренутни резултат изражен у броју победа рунди за оба тима. Поред утицаја појединачних акција играча, на исход меча утичу и одлуке на нивоу тима. Рад издваја ослањање на тимске одлуке у имплементацији модела. У раду су имплементирана три модела помоћу следећих методологија: логистичка регресија, *XGBoost* и неуронске мреже. За улаз се користе информације о стању меча на почетку сваке рунде. За скуп података коришћено је 6500 мечева од 1. априла 2020. до 20. априла 2021. године. Из рада се закључује да су неуронска мрежа и *XGBoost* оптималнија решења у односу на регресију, док између резултата ове две методологије нема разлика. Рад [1] сугерише да за реализацију модела у пројекту треба користити неуронску мрежу или *XGBoost*. Међутим, рад не урачунава појединачне перформансе и ранг играча у исходу меча. Такође не укључује ни рангирање тимова, историју надметања између истих и утицај игране мапе на сам тим.

Рад [2] анализира утицај државе из које долази играч на његове перформансе у такмичарским видео играма. Рад је општији у смислу да обухвата више различитих такмичарских видео игара, али се закључак може применити и на проблему предикције исхода CS:GO меча. Анализирани су односи традиције државе у игрању *Esports* игара, карактеристика државе и перформанси у играма. Коришћена је регресиона и корелациона анализа. Скуп података се састоји из 500 најбољих награда за играче на *ESports* турнирима у периоду од 2004. до 2014. године. Поред тога, коришћени су подаци о карактеристикама државе: Хофстедове културне димензије [3], Светско истраживање вредности [4] и Извештај о глобалној конкурентности [5]. Регресиона анализа се користила за процену утицаја фактора новца, а корелациона анализа је коришћена за процену утицаја карактеристика државе. У раду је показано да држава играча утиче на перформансе у *Esports* играма, као и да се фактор државе више испољава на вишим класама турнира у односу на ниже. У оквиру рада, нису узете у обзир индивидуалне карактеристике играча.

Рад [6] анализира предикцију меча на основу статистике сваког појединачног играча. Статистика играча обухвата различите податке о резултату које је играч постигао у прошлим мечевима. Акценат је на одабиру података које је потребно укључити како би се играчи груписали на основу њиховог стила игре помоћу *KMeans* кластеровања. Kоришћени алгоритми су *KMeans* за кластеровање и *Random Forest* модел за предикцију исхода меча. Поред избора података, истраживање се базира и на погодном избору алгоритама за наведене податке, одабир кластера за *KMeans* алгоритам, као и одабир алгоритма неуронске мреже који ће довести до најбољих резултата. Рад ће нам помоћи како би направили што релевантнији скуп података и како би на основу тог скупа направили модел који ће што боље репрезентовати ‘снагу’ тимова.

**3. МЕТОД**

У наредним поглављима изложени су скуп података, начин на који су креирани модели и спроведени експерименти.

**3.1. Скуп података**

Скуп података [7] који је коришћен за имплементацију овог система је преузет са интернета и јавно је доступан. Дати скуп садржи информације о играчима и мечевима за период од 11/2015 до 03/2020 године. Скуп података се састоји из података о мечевима који су подељени у четири фајла, од којих су следећа два од интереса:

* *players.csv* – индивидуалне перформансе играча на свакој мапи за сваки меч
* *results.csv* – информације о исходима мечева по мапама

На самом почетку, вршено је претпроцесирање скупа података. Коришћене су различите технике филтрирања и трансформације података. Текстуални подаци су техником *label encoding*-а, претварани у нумеричке вредности, које су затим скалиране како не би дошло до постављања лажног приоритета међу обележјима. Скуп података је подељен на тренинг и тест скуп у односу 80:20 у корист тренинг скупа. С обзиром на то да је након филтрирања добијен скуп од преко 75 000 мечева, сматрано је да је 20% података сасвим довољно за тестни скуп.

**3.2. Креирање модела за предикцију исхода меча**

Идеја овог рада је била да се испроба неколико модела који се помињу у претходно наведеној литератури и да се одреди најпогоднији за конкретан проблем решаван у раду. Испробани су следећи модели (алгоритми):

* *Random Forest Regressor*
* *XGBoost Regressor*
* Вештачка неуронска мрежа

Као улаз у моделе користио се иницијални скуп претпроцесираних података, којима су касније додавани подаци у зависности од експеримента. *Random Forest Regressor* и *XGBoost Regressor* су испробани по угледу на радове анализиране у поглављу 2. Такође, креирана је и вештачка неуронска мрежа по угледу на рад [1]. Мрежа се састоји из два слоја, где први слој за активациону функцију има *ReLU,* док други има *sigmoid.*

**3.3. Испитивање утицаја различитих података на исход меча**

У претходном поглављу представљен је иницијални експеримент који је за циљ имао да одреди најпогоднији модел за овакав систем. Након тога уследили су експерименти где се испитивао утицај различитих података на исход меча.

Прво је анализиран податак о стилу игре играча. Податак о стилу игре добијен је кластеровањем играча на основу његових статистичких података током каријере. Коришћен је *KMeans* алгоритам за кластеровање. Идеја која стоји иза коришћења *KMeans* алгоритма и уврштавања податка о стилу игре садржана је у литератури [6]. Испробан је различити број кластера: 3, 5, 7, 9. Утврђено је да је оптималан број клатера 7.

Затим, следећи експеримент је за циљ имао да испита утицај и процени формулу за рачунање *rating*-а играча. Податак о *rating*-у играча је садржан у скупу података (обележје *rating 2.0*). Међутим, формула за рачунање овог рејтинга није јавно доступна. Стога вреди испитати да ли се *rating 2.0* може апроксимирати из осталих података о играчу да би се модел могао применити и на играче за које ова оцена није доступна. На слици 1 су приказани подаци који утичу на *rating 2.0* [8].



Слика 1. Дијаграм података који утичу на *rating 2.0*

На основу истраживања и коришћења додатних података из скупа креирана је формула за процену *rating*-а играча:

$$CustomRating=0.0073\*KAST+0.3591\*KPR+\left(-0.5329\*DPR\right)+0.2372\*Impact+0.0032\*ADR+0.1587+0.08\*^{HS}/\_{numberOfRoundsInMatch}+0.8\* ^{FKDIFF}/\_{numberOfRoundsInMatch}$$

Коначно, испитиван је утицај фактора државе играча на исход меча. Мотивисано радом [2], одлучено је да се испроба да ли ће податак о држави играча позитивно или негативно утицати на предикцију исхода меча. Самостално су осмишљене следеће метрике везане за државу:

1. Информација о држави – Свакој држави се доделио нумерички идентификатор, који је касније скалиран.
2. Просечна вредност рејтинга свих играча из државе посматраног играча. За сваког играча у мечу користила се по једна вредност.
3. Степен поклапања државе у оквиру једног тима, која преставља удео одређене државе у оквиру тима. Рачуна се по формули:

$$\frac{\sum\_{1}^{brd}broj\\_igrača\\_iz\\_države\\_u\\_timu^{2}}{broj\\_igrača\\_u\\_timu^{2}},$$

где *brd* представља број различитих држава играча у оквиру једног тима.

1. Удео броја играча из државе посматраног играча у укупном броју играча у посматраном скупу података.
2. Просечан рејтинг државе, на нивоу тима. Узима се податак из тачке 2, с тим да овог пута за сваки тим постоји по једна вредност. Овим је покушана да се смањи димензионалност.

**4. РЕЗУЛТАТИ И ДИСКУСИЈА**

Иницијални експеримент је за улаз у модел садржао статистичке информације играча на основу историје играча, при чему је важно напоменути да је за *rating* играча коришћен просечан *rating* играча у целокупној каријери како би се избегао привидно добар резултат. Он је као резултате дао тачност од 62.16% за *Random Forest,* 59.51% ѕа *XGBoost* и 58.9% за неуронску мрежу.

Други експеримент, који је испитивао како податка о стилу игре играча утиче на исход меча, дао је најбољу тачност када се играчи поделе у односу на седам стилова игре (који одговарају седам кластера добијеним *KMeans* алгоритмом). Тачност која је постигнута за седам кластера је 60.03% за *XGBoost,* 62.23% за *Random Forest* и 58.9% за неуронску мрежу.

Трећи експеримент, који је испитивао утицај израчунатог *rating*-а играча, чија је формула дата у претходном поглављу, дао је мала побољшања у односу на податак о *rating-*у из скупа података. Постигнута је тачност од 62.89% за *XGBoost,* 61.14% за *Random Forest* и 60.44% за неуронску мрежу.

У четвртом експерименту је испитан утицај фактора државе на предикцију исхода меча. У претходном поглављу приказане су испробане метрике. Најбоље се показала метрика из тачке 2, за коју се добила тачност од 62.36% за *Random Forest,* 59.99% за *XGBoost* и 59.2% за неуронску мрежу. При смишљању метрика везаних за државу, очекивано је било да метрика из тачке 3 (степен поклапања држава играча у оквиру тима) да најбоље резултате, што није био случај.

Из приложених резултата експеримената, може се закључити да се у општем случају најбоље показао *Random Forest* алгоритам. Испробана је и комбина­ци­ја претходних експеримената на *Random Forest* алго­ритму, где су узете метрике са најбољим решењима из претходних експеримената. Узет је фактор државе са метриком 2 и подела у седам кластера према стилу игре. За рејтинг су, као поређење, узети посебни слу­чајеви за просечан рејтинг из скупа података и изра­чунат рејтинг. Резултати овог покушаја су приказани у табели 1.

|  |  |  |
| --- | --- | --- |
| Фактор државе 2 и к = 7 | Просечан *rating* | Израчунат *rating* |
| *Random Forest* | 62.623 | **62.790** |

Табела 1. Финални резултати

Може се закључити, на основу претходних експери­мената, да највећи утицај од тестираних параметара имају фактор државе са метриком 2 и *KMeans* алго­ритам за 7 кластера са којим је остварена тачност од 62.79%.

**5. ЗАКЉУЧАК**

У овом раду је решаван проблема предикције исхода професионалних тимских мечева у игри CS:GO. За потребе овог рада коришћен је јавно доступан скуп података[7]. С обзиром на природу самог проблема и на присутност људског фактора, био сам мотивисан да истражим шта све може утицати на исход једног професионалног меча. Кроз неколико експеримената је покушано да се пронађе што боље решење за дати проблем. Испробани су *Random Forest*, *XGBoost* алгоритам и вештачка неуронска мрежа као решење проблема. Утврђено је да се најбоље показао *Random Forest Regressor* са тачности од 62.79%.

У наставку ће бити дати предлози за даља унапређења овог решења:

* Коришћење и парсирање демо фајлова меча, где би се могли извући детаљнији подаци и потенцијално побољшати резултати.
* Испитивање повећања и пада перформанси играча у односу на доба године, где би се могло закључити да ли је играч у успону или је већ достигао врхунац своје каријере.
* Решење имплементирано у овом раду узима статистику из целог скупа података не давајући већу тежину новијим подацима. Као проширење се може увести и давање већег значаја новијим подацима везаних за статистику играча јер она боље осликава тренутно стање.
* Испробавање других, комплекснијих модела би могло дати још већу тачност.

**6. ЛИТЕРАТУРА**

[1] Peter Xenopoulos, Bruno Coelho, Claudio Silva - New York University, “Optimal Team Economic Decisions in Counter-Strike”

[2] Petr Parshakov, Marina Zavertiaeva, International Laboratory of Intangible-driven Economy, National Research University Higher School of Economics (2015), “Success in eSports: Does Country Matter?”

[3] Geert Hofstede, “Hofstede's cultural dimensions theory”

[4] Christian Haerpfer, Alejandro Moreno, Christian Welzel, Bi Puranen,  Alejandro Moreno “World values survey”

[5] World Economic Forum – “Global Competitiveness Report”

[6] Student Zachary Schmidt, Dr. Mike Preuss, Dr. Rens Meerhof (2020) - Leiden Institute of Advanced Computer Science, “Esports Match Result Prediction for a Decision Support System in Counter-Strike: Global Offensive”

[7] Коришћени скуп података – Kaggle вебсајт, Mateus Dauernheimer Machado, “CS:GO Professional Matches Analysis”

[8] *Rating 2.0,* <https://www.hltv.org/news/20695/introducing-rating-20> [приступљено: 12.09.2022.]

**Кратка биографија:**

|  |  |
| --- | --- |
|  | **Душан Петровић** рођен је 1998. године у Београду. Основне академске студије завршио је 2021. године на Факултету техничих наука у Новом Саду, на ком брани и мастер рад 2022. године из области **Електротехнике и рачунарства** контакт: dule\_srb@hotmail.rs |