

# Využití strojového učení na trhu kurzového sázení ve sportovní oblasti

Bc. Radim Hajda

---

Diplomová práce  
2021



Univerzita Tomáše Bati ve Zlíně  
Fakulta aplikované informatiky

---

Univerzita Tomáše Bati ve Zlíně  
Fakulta aplikované informatiky  
Ústav počítačových a komunikačních systémů

Akademický rok: 2020/2021

## ZADÁNÍ DIPLOMOVÉ PRÁCE (projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: Bc. Radim Hajda  
Osobní číslo: A18560  
Studijní program: N3902 Inženýrská informatika  
Studijní obor: Počítačové a komunikační systémy  
Forma studia: Prezenční  
Téma práce: Využití strojového učení na trhu kurzového sázení ve sportovní oblasti  
Téma práce anglicky: Using Machine Learning at the Betting Market in the Sport Field

### Zásady pro vypracování

1. Seznamte se s oblastí kurzového sázení.
2. Seznamte se s oblastí strojového učení.
3. Vyberte vhodnou metodu pro predikování výsledků sportovních utkání a implementujte ji.
4. Vytvořte modely pro predikování výsledků sportovních utkání.
5. Srovnajte a zhodnoťte dosažené výsledky.

Forma zpracování diplomové práce: **Tištěná/elektronická**

**Seznam doporučené literatury:**

1. GOODFELLOW, Ian, Yoshua BENGIO a Aaron COURVILLE. *Deep learning*. Cambridge, Massachusetts: The MIT Press, [2016], xxii, 775 s. Adaptive computation and machine learning. ISBN 978-0-262-03561-3.
2. RASCHKA, Sebastian a Vahid MIRJALILI. *Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow*. Second edition. Birmingham: Packt, 2017, xviii, 595 s. ISBN 978-1-78712-593-3.
3. CHOLLET, François. *Deep learning v jazyku Python: knihovny Keras, Tensorflow*. Praha: Grada Publishing, 2019, 328 s. Knihovna programátora. ISBN 978-80-247-3100-1.
4. BURKOV, Andriy. *The Hundred-Page Machine Learning Book*. 2019. ISBN 978-1999579517.
5. RASCHKA, Sebastian a Vahid MIRJALILI. *Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow*. Second edition. Birmingham: Packt, 2017, xviii, 595 s. ISBN 9781787125933.
6. ALPAYDIN, Ethem. *Introduction to machine learning*. 2nd ed. Cambridge, Massachusetts: MIT Press, c2010, xl, 537 s. Adaptive computation and machine learning. ISBN 9780262012430.
7. JONES, M. Tim. *Artificial intelligence: a systems approach*. Sudbury: Jones and Bartlett Publishers, 2008, 498 s. Computer science series. ISBN 9780763773373.

Vedoucí diplomové práce: **doc. Ing. Zuzana Komínková Oplatková, Ph.D.**  
Ústav informatiky a umělé inteligence

Datum zadání diplomové práce: **15. ledna 2021**  
Termín odevzdání diplomové práce: **17. května 2021**

**doc. Mgr. Milan Adámek, Ph.D. v.r.**  
děkan



**Ing. Miroslav Matýšek, Ph.D. v.r.**  
ředitel ústavu

Ve Zlíně dne 15. ledna 2021

**Prohlašuji, že**

- beru na vědomí, že odevzdáním diplomové/bakalářské práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, bez ohledu na výsledek obhajoby;
- beru na vědomí, že diplomová/bakalářská práce bude uložena v elektronické podobě v univerzitním informačním systému dostupná k prezenčnímu nahlédnutí, že jeden výtisk diplomové/bakalářské práce bude uložen v příruční knihovně Fakulty aplikované informatiky Univerzity Tomáše Bati ve Zlíně a jeden výtisk bude uložen u vedoucího práce;
- byl/a jsem seznámen/a s tím, že na moji diplomovou/bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- beru na vědomí, že podle § 60 odst. 1 autorského zákona má UTB ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- beru na vědomí, že podle § 60 odst. 2 a 3 autorského zákona mohu užít své dílo – diplomovou/bakalářskou práci nebo poskytnout licenci k jejímu využití jen připouští-li tak licenční smlouva uzavřená mezi mnou a Univerzitou Tomáše Bati ve Zlíně s tím, že vyrovnání případného přiměřeného příspěvku na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše) bude rovněž předmětem této licenční smlouvy;
- beru na vědomí, že pokud bylo k vypracování diplomové/bakalářské práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tedy pouze k nekomerčnímu využití), nelze výsledky diplomové práce využít ke komerčním účelům;
- beru na vědomí, že pokud je výstupem diplomové/bakalářské práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá. Neodevzdání této součásti může být důvodem k neobhájení práce.

**Prohlašuji,**

- že jsem na diplomové práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.
- že odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

Ve Zlíně, dne  
26. 05. 2021

Radim Hajda v. r.  
podpis diplomanta

## **ABSTRAKT**

Cílem této práce je profitovat na živém fotbalovém sázkovém trhu pomocí strojového učení. Doposud se akademické práce věnovaly převážně předzápasovým sportovním sázkovým trhům. Pomocí přesnosti se porovnávají vytvořené modely s bookmakerem. Nejlepší model dosahoval o 2,84 % větší přesnosti než bookmaker. Oba navržené modely vykazovaly zisk v sezoně 2020/2021 anglické nejvyšší fotbalové soutěže Premier League.

Klíčová slova: strojové učení, neuronové sítě, živé sázkové trhy, fotbal

## **ABSTRACT**

The goal of this thesis is to generate a profit in in-play betting markets using machine learning. Lots of academic works has focused mainly on pre match sports betting markets. The created models compared with the bookmaker using accuracy. The best model achieved 2,84 % higher accuracy than the bookmaker. Both created models reached a profit in the 2020/2021 season of the English Premier League.

Keywords: machine learning, neural networks, in-play betting markets, football

Rád bych poděkoval vedoucí diplomové práce paní doc. Ing. Zuzaně Komínkové Oplatkové, Ph.D. za odborné vedení, cenné připomínky a čas při řešení dané problematiky. Dále bych chtěl poděkovat rodině za podporu při studiu.

# OBSAH

<b>ÚVOD</b> .....	<b>9</b>
<b>I TEORETICKÁ ČÁST</b> .....	<b>10</b>
<b>1 ÚVOD DO KURZOVÉHO SÁZENÍ</b> .....	<b>11</b>
1.1 SÁZKOVÉ TRHY .....	11
1.2 KURZY .....	11
1.2.1 Desetinný kurz .....	11
1.2.2 Zlomkový kurz .....	11
1.2.3 Americký kurz.....	11
1.3 BOOKMAKER .....	12
1.4 TYPY SÁZEK .....	12
1.4.1 1X2 sázky.....	12
1.4.2 Nad/pod sázky.....	12
1.4.3 Handicapové sázky.....	12
1.4.4 Live sázky .....	12
1.5 BOOKMAKEROVA MARŽE .....	13
1.6 ARBITRÁŽNÍ SÁZENÍ .....	13
<b>2 LITERÁRNÍ REŠERŠE</b> .....	<b>14</b>
2.1 ŽIVÉ FOTBALOVÉ SÁZKOVÉ TRHY.....	14
2.2 PŘEDZÁPASOVÉ FOTBALOVÉ SÁZKOVÉ TRHY .....	14
<b>3 STROJOVÉ UČENÍ (MACHINE LEARNING)</b> .....	<b>15</b>
3.1 TYPY UČENÍ.....	15
3.1.1 Učení s učitelem.....	15
3.1.2 Učení bez učitele .....	16
3.1.3 Kombinace učení s učitelem a bez učitele .....	16
3.1.4 Posilované učení.....	16
3.2 TRÉNOVACÍ, TESTOVACÍ A VALIDAČNÍ MNOŽINA .....	16
3.3 PŘEUČENÍ (OVERFITTING) A PODUČENÍ (UNDERFITTING).....	17
<b>4 UMĚLÉ NEURONOVÉ SÍŤE</b> .....	<b>18</b>
4.1 MATEMATICKÝ MODEL NEURONU .....	18
4.2 AKTIVAČNÍ (PŘENOSOVÁ) FUNKCE.....	19
4.3 ARCHITEKTURY NEURONOVÝCH SÍŤÍ .....	19
4.3.1 Cyklická (rekurentní) síť .....	19
4.4 VÍCEVRSTVÝ PERCEPTRON (MULTI-LAYERED PERCEPTRONS, MLP).....	20
4.5 REGULARIZACE .....	20
4.5.1 Váhová regularizace L1 a L2 .....	20
4.5.2 Výpadek (Dropout) .....	21
4.5.3 Předčasné ukončení (Early Stopping).....	21

5.1	POPIS GENETICKÉHO ALGORITMU .....	22
5.1.1	Inicializace populace .....	22
5.1.2	Ohodnocení účelovou funkcí .....	22
5.1.4	Křížení .....	23
5.1.5	Mutace .....	23
<b>II</b>	<b>PRAKTICKÁ ČÁST .....</b>	<b>24</b>
<b>6</b>	<b>PŘÍPRAVA PRO TRÉNOVÁNÍ A TESTOVÁNÍ NEURONOVÝCH SÍTÍ.....</b>	<b>25</b>
6.1	SBĚR DAT A KURZŮ .....	25
6.1.1	Poločasová statistika .....	25
6.1.3	Poločasové kurzy .....	26
6.3	VÝBĚR ATRIBUTŮ POMOCÍ GENETICKÉHO ALGORITMU .....	27
6.3.1	Popis řešení .....	27
6.3.2	Účelová funkce.....	28
6.4	VÝBĚR SÁZKOVÉ STRATEGIE .....	28
6.4.1	Fixní sázka .....	28
6.4.2	Fixní návratnost.....	28
6.4.3	Kellyho kritérium .....	29
6.5	SÁZKOVÝ SIMULÁTOR .....	30
<b>7</b>	<b>EXPERIMENTY A VÝSLEDKY .....</b>	<b>31</b>
7.1	MODEL I.....	32
7.1.1	Použité nastavení.....	32
7.1.2	Výsledky sázkových strategií.....	33
7.2	MODEL II .....	40
7.2.1	Použité nastavení.....	40
7.2.2	Výsledky sázkových strategií.....	41
<b>8</b>	<b>ZHODNOCENÍ VÝSLEDKŮ .....</b>	<b>51</b>
8.1	POROVNÁNÍ SÁZKOVÝCH STRATEGIÍ .....	51
8.2	POROVNÁNÍ KURZŮ .....	54
8.3	POROVNÁNÍ PŘESNOSTÍ .....	56
	<b>ZÁVĚR .....</b>	<b>58</b>
	<b>SEZNAM POUŽITÉ LITERATURY.....</b>	<b>60</b>
	<b>SEZNAM POUŽITÝCH ZKRATEK.....</b>	<b>62</b>
	<b>SEZNAM OBRÁZKŮ .....</b>	<b>63</b>
	<b>SEZNAM TABULEK.....</b>	<b>64</b>
	<b>SEZNAM PŘÍLOH.....</b>	<b>65</b>



## ÚVOD

Strojové učení našlo ve fotbalovém prostředí uplatnění v několika různorodých oblastech. Pro vytváření statistik v reálném čase fotbalového utkání se používají algoritmy strojového vidění. Strojové vidění našlo uplatnění také v identifikaci osob, které mají z nějakého důvodu zakázaný vstup na stadion nebo jeho prostor. Evropské týmy také využívají strojové učení pro hledání nových hráčů, kteří tým vhodně posílí pro nadcházející sezonu nebo také v průběhu již započaté sezony. Využití ve fotbale nachází i posilované učení, pomocí kterého se učí agenti hrát fotbal, a na základě těchto získaných poznatků se vytváří nové taktiky. Strojové učení má ve fotbalovém světě velký a stále trochu nevyužitý potenciál. Autor práce si myslí, že strojové učení se v následujících letech stane nedílnou součástí systému VAR (video asistent rozhodčího) a pomůže tomuto systému minimalizovat chyby pomocí strojového vidění, tím že systém bude videorozhodčímu předkládat již předvybrané záběry a tím se budou redukovat chybná rozhodnutí, které byly udělány na základě záběru ze špatného úhlu, a hlavně by takový systém redukoval zbytečné prostoje, které dosahují i více než 5 minut. V posledních letech se také algoritmy strojového učení využívají pro predikování nejen výsledků, ale i různých událostí, které nastanou nebo nenastanou ve sportovním utkání. Na základě těchto predikcí poté sázkové kanceláře stanoví kurz, na který si může sázkař vsadit. Tato diplomová práce se orientuje a specializuje na live sázky a zabývá se vytvořením "bookmakera" - modelu vytvářející sázkové kurzy - pomocí neuronových sítí.

Cílem teoretické části práce je seznámit čtenáře se základními pojmy kurzového sázení a nutným teoretickým výkladem z oblasti strojového učení, neuronových sítí a genetických algoritmů, které jsou stěžejní pro pochopení vytvořených modelů, které se nachází v praktické části práce.

Hlavním cílem praktické části práce je demonstrovat možnost porazit bookmakera v jeho hře pomocí neuronových sítí i ve sportovních sázkách při právě probíhajících sportovních utkání. Hlavními kritérii pro vyhodnocení vytvořených modelů je přesnost a ziskovost.

## **I. TEORETICKÁ ČÁST**

## 1 ÚVOD DO KURZOVÉHO SÁZENÍ

Kurzové sázení lze definovat jako hru mezi sázkařem a bookmakerem, při které se oba snaží maximalizovat svůj zisk a minimalizovat ztráty.

### 1.1 Sázkové trhy

Sázkový trh lze považovat za kategorii konkrétního typu sázky. Pro jakoukoliv událost mohou sázkové kanceláře otevřít mnoho různých sázkových trhů [1]. Například pro fotbalový zápas se bude jednat o trhy od vítěze utkání, až po přesný počet branek jednoho z týmů. V dnešní době lze sázet prakticky na jakoukoliv sportovní, ale i společenskou událost.

### 1.2 Kurzy

Existuje několik podob sázkových kurzů. Zatímco se v České republice používá desetinný formát zápisu kurzů, je možné se setkat na zahraničních webech i s jiným zápisem, jako je zlomkový nebo americký.

#### 1.2.1 Desetinný kurz

Desetinný kurz je převrácená hodnota pravděpodobnosti, že ne/nastane nějaká událost, na kterou je možné si vsadit [1]. Například pro pravděpodobnost 0,2 bude vypsán kurz  $0,2^{-1} = 5$ . Samozřejmě je možné stejně přepočítat i kurz 5 na pravděpodobnost  $5^{-1} = 0,2$ .

#### 1.2.2 Zlomkový kurz

Zlomkový zápis kurzů se používá především ve Velké Británii. Například zlomkový kurz 4 / 1 udává, že je zapotřebí vsadit jednu jednotku bankrollu, aby čistý zisk z vyhrané sázky činil 4 jednotky. Pokud by 1 jednotka odpovídala 100 Kč, tak by byl čistý zisk ze sázky 400 Kč. Zlomkový kurz se přepočítá na desetinný tak, že sečteme čitatele a jmenovatele zlomkového kurzu a podělíme ho jmenovatelem, a proto zlomkový kurz 4 / 1 odpovídá desetinnému kurzu  $(4+1) / 1 = 5$  [1].

#### 1.2.3 Americký kurz

Nejvíce zmatečně se pro tuzemské sázkaře můžou jevit americké kurzy, které nabývají kladných, nebo záporných hodnot. Například americký kurz +400 udává, že při sázce 100 dolarů by činil čistý zisk 400 dolarů. Pokud je americký kurz záporný, například -400 znamená to, že musíme vsadit 400 dolarů, aby čistý zisk činil 100 dolarů [1].

### 1.3 Bookmaker

Bookmaker je osoba nebo společnost, která vypisuje a upravuje kurzy na vypsané události, ale taky přijímá sázky a vyplácí výhry [1].

### 1.4 Typy sázek

Dnes je bookmakerem nabízeno velké množství sázkových příležitostí, proto jsou v této kapitole popsány jen ty nejoblíbenější mezi sázkaři.

#### 1.4.1 1X2 sázky

Jedna z nejstarších a nejpoblíbenějších sázek hlavně ve fotbale. Cílem sázkaře je správně predikovat výsledek utkání, které může skončit třemi různými výsledky, a to výhrou domácího týmu, výhrou hostujícího týmu, nebo remízou [2]. Jedničkou se označuje sázka na vítězství domácího týmu. Dvojka značí sázku na vítězství hostujícího týmu a X je označení pro sázku na remízu. U tuzemských sázkových kanceláří se setkáme s nulou jako označením pro remízu při výběru sázky v nabídce bookmakera.

#### 1.4.2 Nad/pod sázky

U nad/pod sázek se predikuje, zda stanovená hodnota (góly, gemy, sety, body atd.) přesáhne/nepřesáhne hranici stanovenou bookmakerem [2]. Například sázka *Nad 2,5 gólu* ve fotbalovém utkání je výherní, pokud je součet branek v utkání 3, nebo více, a naopak je proherní pokud součet branek obou týmů je 0,1, nebo 2.

#### 1.4.3 Handicapové sázky

U handicapové sázky dochází k přidělení imaginární výhody jednomu z týmů [2]. Například při fotbalovém utkání to znamená, že jeden z týmů před začátkem utkání vede 1:0. Tento typ sázek se nejčastěji využívá při sázení na favorita utkání, který imaginárně prohrává, čímž se dosáhne vyššího kurzu na favorita zápasu.

#### 1.4.4 Live sázky

Online sázkové kanceláře nabízí možnost uzavírat „live“ sázky. Jedná se o speciální typ sázky, kdy si sázkař může vsadit na již probíhající sportovní události. Kurzy se mění v reálném čase a odráží aktuální stav sportovní události [2]. Typickým příkladem je sázka na počet branek ve fotbalovém utkání o poločasové přestávce nebo v jakékoli jiné minutě zápasu.

## 1.5 Bookmakerova marže

Je třeba si uvědomit, že sázkové kanceláře nehrají fér a dochází z jejich strany k podhodnocování kurzů [1]. Jako příklad této techniky je uveden hod mincí. Jelikož je pravděpodobnost pro oba stavy 50%, bylo by fér, kdyby vypsané kurzy byly 2,0 : 2,0, ale ve skutečnosti bookmaker vypíše kurzy 1,9 : 1,9. V ukázkovém příkladě je tedy bookmakerova marže  $(1,9^{-1} + 1,9^{-1} - 1) * 100 = 5,2\%$ .

## 1.6 Arbitrážní sázení

Arbitrážní sázení je metoda, která garantuje nízký zisk bez rizika prohry. Jedná se o využití rozdílů kurzů mezi sázkovými kancelářemi [1]. Názorný příklad je uveden na sázce *Nad/Pod 2,5 branek* ve fotbalovém utkání (Tabulka 1).

Tabulka 1 – Ukázka arbitrážního sázení

	Nad 2,5 branek	Pod 2,5 branek	Marže [%]
Kurz bookmakera A	1,60	2,40	5,98
Kurz bookmakera B	1,80	2,10	4,76
Arbitráž	1,80	2,40	-2,78

V modelovém příkladu podle tabulky (Tabulka 1) by došlo k vsazení sázky *Nad 2,5 branek* za 100 Kč u bookmakera B v kurzu 1,8 a zároveň uzavření sázky *Pod 2,5 branek* za 75 Kč u bookmakera A v kurzu 2,4. Celkový vklad by činil 175 Kč. Při takhle přiřazených sázkách je jedno jak utkání dopadne, protože ať už výsledek utkání bude jakýkoliv v obou případech nám jeden z bookmakerů vyplatí 180 Kč, takže čistý zisk z téhle sázky činí 5 Kč.

Na první pohled se jeví arbitrážní sázení jako ideální příležitost pro sázení bez rizika. Realita je dost odlišná, jelikož je třeba překonat několik nástrah.

Sázkové kanceláře sledují konkurenci a své kurzy neustále upravují. Šance pro využití arbitráže je tedy otázkou několika sekund než minut. Sázkové kanceláře ale také mohou sázku odmítnout, nebo limitovat výši vkladu. Také je zapotřebí mít vytvořeno několik účtů u různých sázkových kanceláří a neustále přelévat mezi nimi finanční prostředky na sázení, protože výhra u jedné, znamená automaticky prohru u druhé sázkové kanceláře. Pro člověka, který by tyto nástrahy překonal by podle Egona, Verbeeka a Nüesche [3] činil průměrný výnos ze sázky pouze 1,2 %.

## 2 LITERÁRNÍ REŠERŠE

Literární rešerše je rozdělena do dvou podkapitol. První podkapitola rešerše se věnuje pracím, které se zabývají živými fotbalovými sázkovými trhy a ve druhé podkapitole jsou popsány práce, které se zabývají předzápasovými sázkovými fotbalovými trhy. Vzhledem k tomu, že v diplomové práci budou využívány neuronové sítě, literární rešerše byla zaměřena především na používání neuronových sítí v jednotlivých typech sázkových trhů.

### 2.1 Živé fotbalové sázkové trhy

Pettersson a Nyquist [4] používají rekurentní neuronové sítě (RNN) pro predikování vítěze fotbalového utkání. Věnují se jak predikcím před začátkem utkání, tak i předpovědím po každých odehraných 15 minutách zápasu.

### 2.2 Předzápasové fotbalové sázkové trhy

Borøy-Johnsen [5] se zabývá porovnáním několika modelů neuronových sítí, které predikují vítěze fotbalového utkání v anglické Premier League a používá atributy od hodnocení jednotlivých hráčů až po týmové statistiky. Srovnává také různé sázkařské strategie, pomocí kterých generoval zisk, takže se mu podařilo demonstrovat, že neuronové sítě jsou vhodným nástrojem k poražení bookmakera v jeho vlastní hře.

Samba [6] se ve své práci zabývá použitím hlubokého učení pro predikování vítěze fotbalového zápasu. Oproti Borøy-Johnsenovi se jeho datová sada neskládá pouze z jedné fotbalové soutěže, ale z několika vybraných fotbalových lig. Používá hlavně týmové atributy, ale zavádí i neutrální atributy, které se týkají rozhodčího nebo časové vzdálenosti mezi zápasy. Porovnává různá nastavení neuronových sítí, která se liší hlavně počtem neuronů ve skrytých vrstvách nebo počtem skrytých vrstev.

Tax [7] se zaměřuje na nizozemskou fotbalovou nejvyšší soutěž Eridivisii. Oproti předchozím autorům definuje nové atributy. K nejzajímavějším atributům patří informace o postupu z nižší soutěže, změna manažera, zranění nejlepšího střelce týmu. Autor demonstruje, že pomocí veřejně dostupných dat je možné generovat zisk na vybraných sázkových trzích.

Godin a spol [8] razí úplně jiný přístup a snaží se využít kolektivní znalosti pomocí analýzy mikropostů ze sociální sítě Twitter k predikci výsledků anglické nejvyšší fotbalové ligy Premier League.

### 3 STROJOVÉ UČENÍ (MACHINE LEARNING)

Strojové učení je podobor umělé inteligence, která se zabývá algoritmy pro vytváření modelů naučených na vhodných datech. Data jsou podstatnou součástí strojového učení a mohou pocházet z různých zdrojů – z přírody, ručně vytvořeny lidmi, snímány z různých snímačů a senzorů nebo i generovány jiným algoritmem [9]. U strojového učení se vkládají do algoritmu data, ale i odpovědi, které se očekávají u zpracovaných dat. Výsledkem jsou pravidla. Tato pravidla se aplikují na nová data k vytváření originálních odpovědí [10].



Obrázek 1 – Strojové učení jako programové paradigma [10]

#### 3.1 Typy učení

Mezi základní typy učení patří:

- Učení s učitelem (supervised learning)
- Učení bez učitele (unsupervised learning)
- Kombinace učení s učitelem a bez učitele (semi-supervised learning)
- Posilované učení (reinforcement learning)

dle [11].

##### 3.1.1 Učení s učitelem

Učení s učitelem je jedním z nejběžnějších případů. Jedná se o mapování vstupních dat na známé výstupy, na základě datového souboru příkladů [10]. Typickou úlohou pro učení s učitelem je klasifikační nebo regresní úloha. Dobrým příkladem klasifikace je spamový filtr. Je natrénován na mnoha příkladech e-mailů, které mají označen výstup (spam nebo ham) a musí se naučit, jak klasifikovat nové e-maily. Dalším typickým příkladem je regresní úloha, tedy predikování výstupní číselné hodnoty, jako je například cena automobilu vzhledem k vstupním atributům (počet najetých kilometrů, stáří vozidla, značka). Pro

trénování modelů řešící dané úlohy je zapotřebí velké množství označených dat (např. zmiňovaných automobilů či emailů) [11].

### 3.1.2 Učení bez učitele

Toto odvětví strojového učení se skládá z hledání zajímavých transformací vstupních dat bez pomoci jakéhokoli označeného výstupu. Mezi úlohy učení bez učitele patří datové vizualizace, datové komprese, odstranění šumu nebo lepší pochopení korelace přítomné v daných datech. Mezi dobře známé kategorie učení bez učitele patří redukce dimenzionality a shlukování [9, 10].

### 3.1.3 Kombinace učení s učitelem a bez učitele

Při kombinaci učení s učitelem a bez učitele obsahuje datová sada jak označené výstupy, tak výstupy neoznačené. Počet neoznačených příkladů je obvykle mnohem větší, než počet příkladů označených. Cíl takového učení je stejný jako při učení bez učitele. Předpokládá se, že při použití mnoha neoznačených výstupů s několika vstupy označenými, bude mít model do začátku „tahák“, který pomůže vytvořit lepší model [9].

### 3.1.4 Posilované učení

Při posilovaném učení, agent získává informace o prostředí a učí se vybrat akci, která bude maximalizovat nějakou odměnu. Například neuronová síť, která se „dívá“ na obrazovku videohry a provádí herní akce, aby maximalizovala své herní skóre, může být natrénována pomocí posilovaného učení [10].

## 3.2 Trénovací, testovací a validační množina

Pro vyhodnocení modelu se většinou data rozdělí do tří množin (trénovací, testovací a validační). Pro natrénování modelu slouží trénovací množina. Pro vyladění modelu a testování různých nastavení parametrů a konfigurací se používá validační množina. Po vyladění modelu na validační množině se testuje na množině testovací. Doporučený poměr dat v množinách se udává 2:1:1. Při výběru dat do množin je třeba si dát pozor, aby všechny třídy v každé množině byly stejně zastoupeny. Například snažíme-li se klasifikovat číslice 0-9, testovací množina by neměla obsahovat pouze číslice od 0 do 7 a validační a testovací množina pouze číslice 8 a 9.



### 3.3 Přeučení (overfitting) a podučení (underfitting)

Přeučení je jev, kdy model vykazuje výborné výsledky na trénovací sadě dat, ale selhává na testovací (validační) datové sadě. Nejčastějšími důvody, proč k přeučení dochází, jsou podle [9]:

- Model je pro dostupná data příliš složitý (příliš velký počet vrstev nebo velký počet neuronů v jednotlivých vrstvách)
- Velký počet atributů, ale málo příkladů v trénovací sadě

Podučení je neschopnost modelu vykazovat dobré výsledky ani na trénovací datové sadě [12]. Nejčastějšími důvody, proč k podučení dochází, jsou podle [9]:

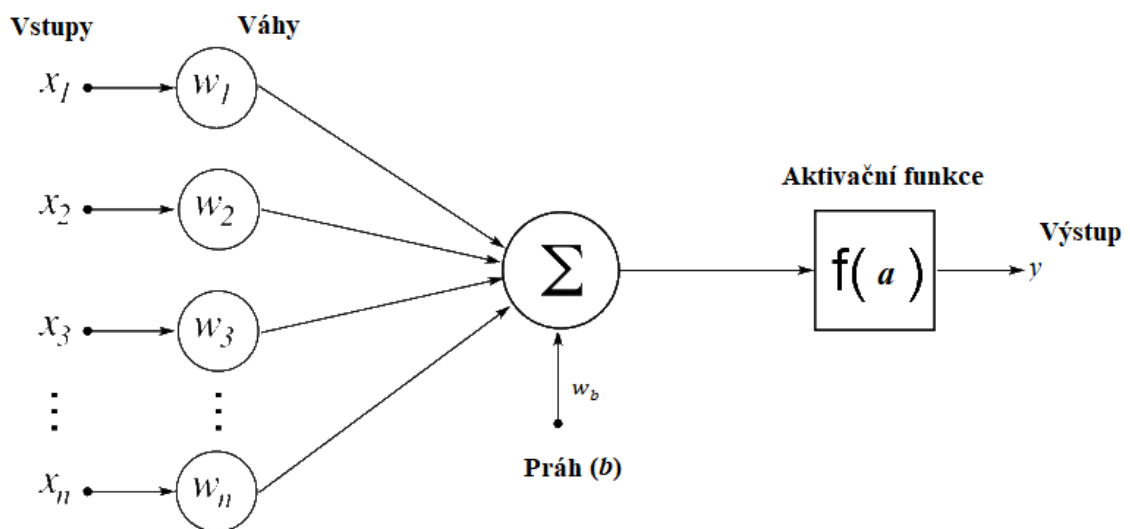
- Model je pro dostupná data příliš jednoduchý
- Navržené atributy nejsou dostatečně informativní

## 4 UMĚLÉ NEURONOVÉ SÍTĚ

Umělé neuronové sítě jsou podobou strojového učení. Jedná se o matematické prediktivní modely, které jsou inspirovány lidským mozkem. Skládají se ze vzájemně propojených neuronů, které jsou z pravidla uspořádané do vrstev [10].

### 4.1 Matematický model neuronu

Umělý neuron se skládá ze vstupů  $x_1$  až  $x_n$  a jednoho výstupu  $y$ . Jednotlivé vstupy jsou ohodnoceny váhami  $w_1$  až  $w_n$ .



Obrázek 2 – Matematický model neuronu [13]

Na základě těchto vah mohou být vstupy zvýhodněny, nebo úplně potlačeny. Pomocí sumace vznikne vnitřní potenciál  $a$  (4.1).

$$a = \sum_{i=1}^n x_i w_i + b w_b \quad (4.1)$$

Vnitřní potenciál je argumentem aktivační (přenosové) funkce.

Pro výstup z přenosové (aktivační) funkce  $f(a)$  tedy platí (4.2).

$$f(a) = TF \left( \sum_{i=1}^n x_i w_i + b w_b \right) \quad (4.2)$$

Kde  $TF$  je libovolná aktivační (přenosová) funkce.

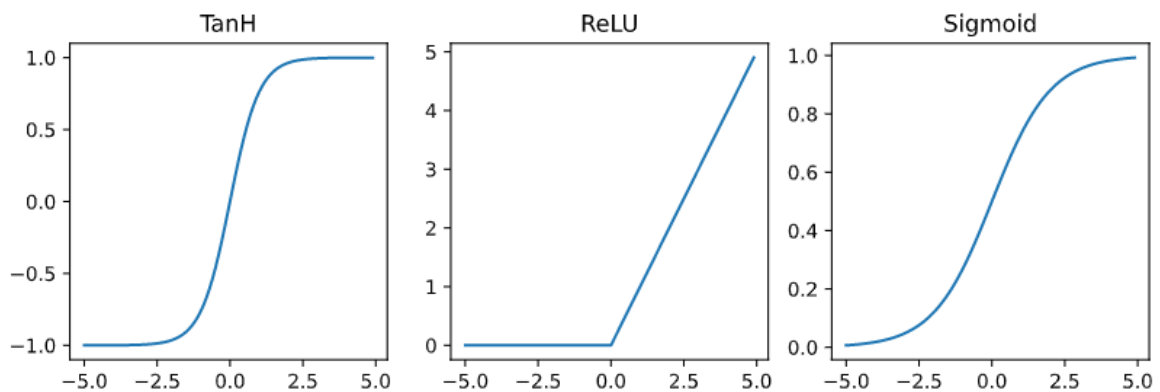
Výstup z aktivační (přenosové) funkce je výstupem sítě  $y$  nebo vstupem do dalšího neuronu. [10,11,12]

## 4.2 Aktivační (přenosová) funkce

Bez aktivačních funkcí, by se neuronové sítě skládaly pouze z lineárních operací, a proto by se jednotlivé vrstvy mohly naučit pouze lineární transformace vstupních dat [10].

Nelineární transformací dat do nového prostoru se může stát, že problém klasifikace, který nebyl lineárně oddělitelný (neřešitelný lineárním klasifikátorem), se stane oddělitelným [14].

Na obrázku (Obrázek 3) jsou zobrazeny pouze vybrané aktivační funkce.



Obrázek 3 – Aktivační funkce

V současné době se doporučuje používat jako aktivační funkci pro neurony ve skryté vrstvě ReLU, případně její modifikace [14].

## 4.3 Architektury neuronových sítí

Rozlišují se dva základní typy architektury: cyklická a acyklická síť.

### 4.3.1 Cyklická (rekurentní) síť

Pro cyklickou síť je podmínkou existence cyklu. To znamená, že výstup z prvního neuronu je vstupem do druhého neuronu, jehož výstup je vstupem třetího neuronu atd, až výstup posledního neuronu je vstupem do prvního neuronu. Nejjednodušším příkladem cyklu je zpětná vazba neuronu, jehož výstup je zároveň jeho vstupem [15].

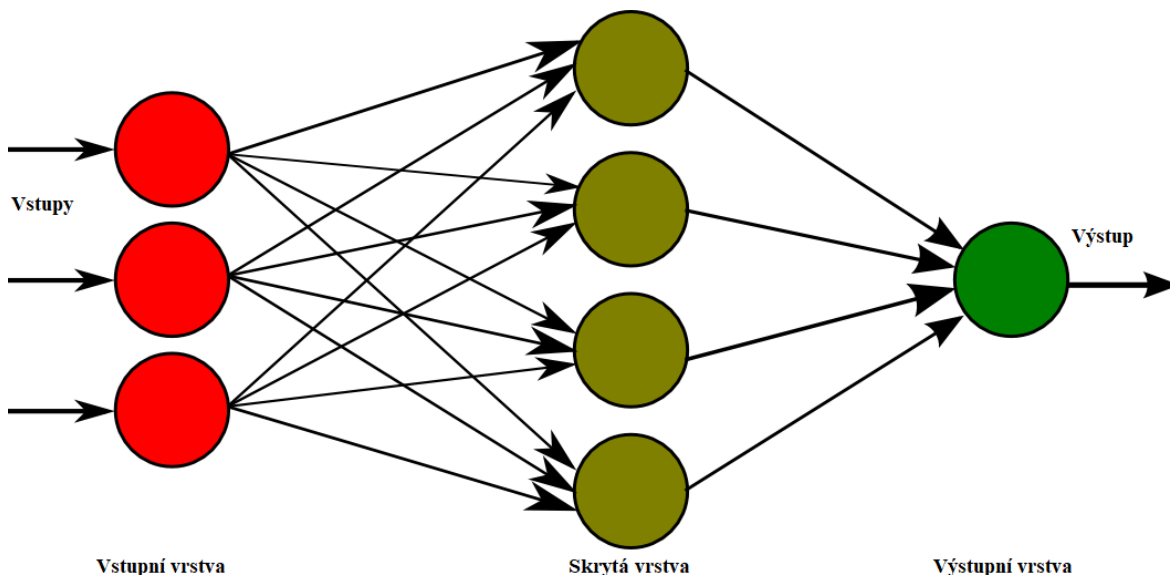
### 4.3.2 Acyklická (dopředná) síť

V acyklických sítích žádný cyklus neexistuje a všechny cesty vedou jedním směrem.

Neurony je vždy možné umístit do vrstev [15].

## 4.4 Vícevrstvý perceptron (Multi-layered Perceptrons, MLP)

Vícevrstvý perceptron (Obrázek 4) se skládá z jedné vstupní vrstvy, jedné nebo více skrytých vrstev a jedné výstupní vrstvy. Každá vrstva je plně propojená s následující vrstvou [11].



Obrázek 4 – Schéma vícevrstvého perceptronu [16]

Vstupní vrstva slouží pouze k předání dat ze vstupů (neurony ve vstupní vrstvě nemají aktivační funkci). Skrytá vrstva má za úkol provést silnou nelineární transformaci pomocí aktivačních funkcí. Výstupní vrstva vrací pomocí aktivační funkce požadovaný výstup. Pro učení MLP se převážně používá učící algoritmus zpětného šíření (backpropagation), ale existují i jiné algoritmy. MLP se používá jak pro klasifikační, tak i pro regresní problémy, záleží pouze na aktivační funkci ve výstupní vrstvě.

## 4.5 Regularizace

Regularizace je pojem, který zahrnuje metody, které nutí učící algoritmus k vytvoření méně složitého modelu a tím vyhnout se přeučení [9]. V podkapitolách jsou popsány pouze nejpoužívanější techniky.

### 4.5.1 Váhová regularizace L1 a L2

L1 regularizace přidává ke ztrátové funkci penalizaci, která je rovna absolutní hodnotě váhových koeficientů. Při L2 regularizaci se ke ztrátové funkci přidává penalizace v podobě druhé mocniny váhových koeficientů. Hlavním rozdílem je, že při L1 regularizaci dojde k vynulování méně důležitých atributů, a proto se L1 regularizace používá i pro výběr

klíčových atributů v případě, že jich máme dostupných velké množství a chceme snížit složitost modelu [14].

#### **4.5.2 Výpadek (Dropout)**

Jedná se o jednu z nejefektivnějších a nejpoužívanějších regularizačních technik [10]. Výpadek se aplikuje na vrstvu a skládá se z náhodného vynechání neuronů z výpočtu, při každém průchodu trénovacího příkladu sítí [9].

#### **4.5.3 Předčasné ukončení (Early Stopping)**

Jedná se o přerušení trénovacího procesu, když nedochází ke zlepšování ztrátové funkce na validační datové sadě [10].

## 5 GENETICKÉ ALGORITMY

Genetické algoritmy jsou založeny na Darwinově teorii přirozeného výběru. Jedná se o heuristiky, které modifikují jedince v populaci za účelem jeho zlepšení [15]. V následující kapitole jsou popsány pouze části Genetického algoritmu, které se používají v praktické části práce pro výběr podmnožiny atributů.

### 5.1 Popis Genetického algoritmu

Genetický algoritmus se skládá z 5 kroků. Nejdříve se vygeneruje náhodná populace, která se skládá z jedinců. Každý jedinec se ohodnotí pomocí účelové funkce. Na základě fitness hodnoty se vyberou vhodní jedinci (rodiče) pro operace křížení a mutaci a tím vzniknou noví jedinci (potomci), kteří vstupují do další generace. Celý proces se opakuje do té doby, než nastane ukončující podmínka [15]. Detailněji je každý krok popsán v jednotlivých podkapitolách.

#### 5.1.1 Inicializace populace

Nejprve se musí náhodně vygenerovat první populace, která je tvořena jedinci. Každý jedinec je tvořen několika parametry, které se nazývají geny. Dohromady geny tvoří chromozom, tedy jednoho jedince populace. Nejčastěji se pro reprezentaci jedince používá binární reprezentace, ale existují i jiné, jako je dekadická nebo symbolická reprezentace [15].

#### 5.1.2 Ohodnocení účelovou funkcí

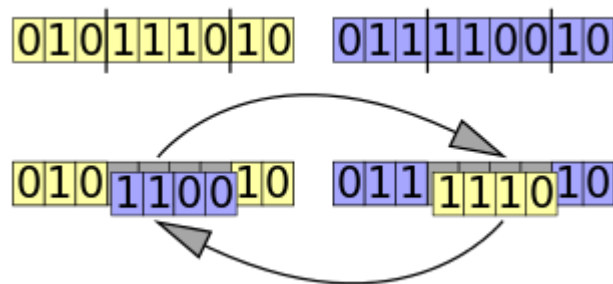
Každý jedinec je ohodnocen pomocí účelové funkce. Hodnota účelové funkce je reálné číslo, které udává vhodnost (fitness) aktuálního řešení pro reprodukci. Hodnota fitness se také normalizuje, aby přímo odpovídala pravděpodobnosti, že jedinec bude vybrán k reprodukci [15].

#### 5.1.3 Selektce

Selektce neboli výběr jedinců pro operace křížení a mutace probíhá na základě hodnoty fitness. Čím větší hodnota fitness, tím větší pravděpodobnost výběru k reprodukci. V praktické části práce je použita turnajová selektce. Při turnajové selekci se vybere  $n$  počet jedinců, kteří spolu soupeří o reprodukci. Vítězem turnaje je jedinec s největší hodnotou fitness. Mezi další selektce, které se používají patří ruletová, nebo pořadová selektce [17].

### 5.1.4 Křížení

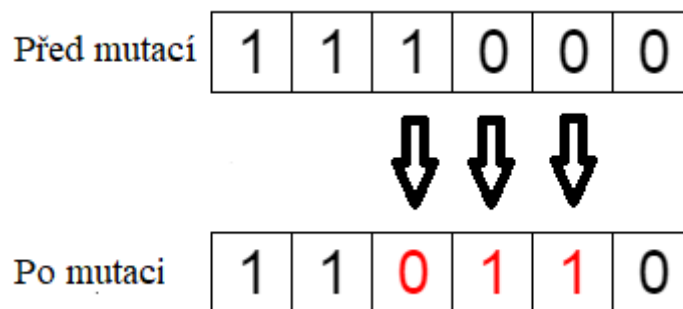
Křížení je nejdůležitější operátor Genetického algoritmu, protože zde dochází k náhodným kombinacím řešení. Křížení probíhá vždy mezi 2 rodiči. Při dvojbodovém křížení se náhodně vyberou 2 body v chromozomu vybraných rodičů a vně těchto bodů dojde k záměně jejich genů (Obrázek 5). Existují, ale i další možnosti křížení, jako je jednobodové, vícebodové, nebo uniformní křížení [17].



Obrázek 5 – Dvoubodové křížení [18]

### 5.1.5 Mutace

Při mutaci se s velmi malou pravděpodobností změní hodnota genu v chromozomu. Z 1 se změní na 0 a obráceně [17].



Obrázek 6 – Ukázka mutace jedince

## **II. PRAKTICKÁ ČÁST**



## 6 PŘÍPRAVA PRO TRÉNOVÁNÍ A TESTOVÁNÍ NEURONOVÝCH SÍTÍ

Před začátkem trénování a testování neuronové sítě je třeba provést řadu nezbytných kroků, aby vytvoření modelu bylo efektivní. V jednotlivých fázích je potřeba se věnovat sběru historických dat a kurzů, výběru vhodného frameworku či vhodné strategie.

### 6.1 Sběr dat a kurzů

Po důkladné rešerši a hledání autor diplomové práce konstatuje, že není dostupná žádná veřejná datová množina, která by obsahovala data již probíhajícího fotbalového utkání, a proto bylo nejnáročnějším úkolem takovou unikátní datovou sadu vytvořit. Některé webové portály poskytují již hotové datasety nebo data formou API za finanční poplatek. Ještě horší situace je ohledně historických kurzů, zatímco se několik webů specializuje na shromažďování kurzů pro předzápasové sázkové trhy, pro živé sázkové trhy nic podobného neexistuje. Oba problémy byly vyřešeny pomocí naprogramovaných skriptů pro sběr dat z vybraných webových stránek. Pro vytvoření skriptů je použit programovací jazyk Python a modul Beautiful Soup [19], který se používá pro usnadnění získávání dat z HTML a XML souborů.

#### 6.1.1 Poločasová statistika

Jako vhodný zdroj dat pro statistiky již probíhajícího utkání byl vybrán fotbalový statistický web **WhoScored.com** [20]. Týmová statistika se provádí zvlášť pro oba zúčastněné týmy. Typickým příkladem je počet střel domácího (hostujícího) týmu na branku nebo počet vstřelených branek domácího (hostujícího) týmu. Pro každý tým bylo z webu WhoScored.com získáno 63 atributů týkajících se pouze statistik po odehraní 1. poločasu utkání.

Shots	Passes	Dribbles	Tackles Attempted	Interceptions	Clearances	Blocks	Offsides	Fouls	Aerial duels	Touches
7   10	236   285	7   12	19   15	4   7	8   9	5   10	0   1	5   4	7   2	343   396


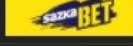

  

Pass Type	Length	Height	Body Parts	Direction	Target Zone
All	All	All	All	All	All
13 Cross	8 28 Long	26 33 Chipped	33 13 Head	7 165 Forward	186 52 Defensive Third
5 Freekick	4 233 Short	277 228 Ground	270 248 Feet	296 95 Backward	115 115 Mid Third
2 Corner	4			133 Left	154 94 Final Third
1 Through Ball	2			126 Right	148
11 Throw In	8				
5 Key Passes	7				

Obrázek 7 – Data pro jednotlivá utkání na webu WhoScored.com

### 6.1.2 Předzápasové kurzy

Pro získání předzápasových kurzů českých sázkových kanceláří na vítěze utkání je použit statistický portál **Livesport.cz** [21]. Výsledné atributy se vypočítají jako průměry kurzů tří sázkových kanceláří (Tipsport, Fortuna a Sazkabet). Kurzy společnosti Chance nejsou započítány, jelikož se jedná o dceřinou společnost sázkové kanceláře Tipsport a tím pádem jsou kurzy na sportovní příležitosti u obou společností totožné.

Kurzy 1X2	H/A	O/U	AH	DŠ	HT/FT	Skóre	L/S	ODG
<b>Základní doba (4)</b>	1. poločas (4)		2. poločas (4)					
Sázková kancelář	1 <sup>▲</sup>		0 <sup>▲</sup>		2 <sup>▲</sup>			
	↑ 2.87	↑ 3.73			↓ 2.46			
	↑ 2.84	↑ 3.75			↓ 2.46			
	↑ 2.90	↑ 3.75			↓ 2.45			
	↑ 2.87	↑ 3.73			↓ 2.46			

Obrázek 8 – Kurzy na Livesport.cz

### 6.1.3 Poločasové kurzy

Aby mohla být vyhodnocena ziskovost modelů, bylo zapotřebí odebírat kurzy o poločasové přestávce fotbalového utkání v reálném čase na webu sázkové kanceláře Tipsport, pro vybranou sázku na *Více rohů v zápasu*.

Více rohů v zápasu				
Manchester U.	5.00	Remíza	7.20	Liverpool 1.30

Obrázek 9 – Live kurzy na Tipsport.cz

## 6.2 Výběr Frameworku

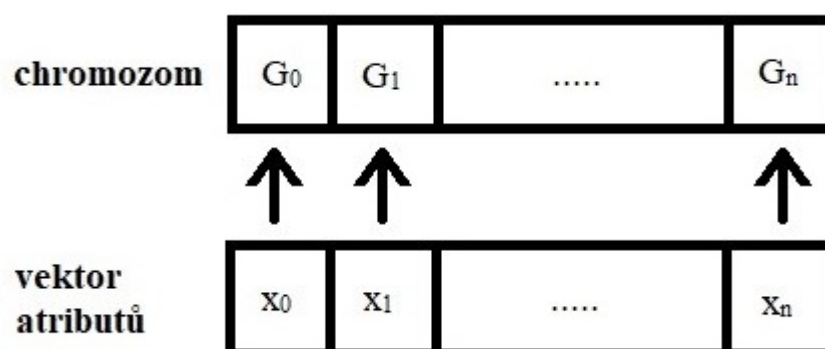
Pro programovací jazyk Python je k dispozici velké množství knihoven pro vytváření neuronových sítí. V praktické části diplomové práce se pro trénování a testování neuronových sítí používá volně dostupné vysokoúrovňové API Keras [22], které využívá na backendu framework TensorFlow [23] od společnosti Google, ale je možné použít i jiné frameworky. Mezi hlavní výhody Kerasu patří jeho jednoduchost, snadná implementace prototypů a velmi dobře zpracovaná dokumentace, která je tvořena převážně ze vzorových příkladů.

## 6.3 Výběr atributů pomocí Genetického algoritmu

Pro výběr atributů pomocí Genetického algoritmu byl pro programovací jazyk Python použit modul **FeatureSelectionGA** [24], u kterého byla vytvořena vlastní třída pro účelovou funkci.

### 6.3.1 Popis řešení

Na začátku algoritmu byla náhodně vygenerována populace, která čítala 50 jedinců (chromozomů). Každý jedinec je tvořen geny, které jsou reprezentovány pomocí binární reprezentace. Pokud gen nabývá hodnoty 1 znamená to, že atribut je vybrán do podmnožiny atributů, které se ohodnotí pomocí účelové funkce, pokud nabývá hodnoty 0 znamená to, že atribut z podmnožiny vynechán, a proto musí být každý atribut namapován na jemu odpovídající gen v chromozomu (Obrázek 9).



Obrázek 10 – Namapování chromozomu na vektor atributů

Pomocí turnajové selekce se vyberou vhodní rodiče k operacím křížení (dvojbodové) a mutace a tím vzniknou 2 potomci do nové generace. Až se naplní nová populace potomky, tak se nahradí stará generace za novou a celý proces se opakuje do té doby, než nastane ukončující podmínka (100 generací). Výsledkem je podmnožina atributů, která dosahuje maximální hodnoty účelové funkce.

### 6.3.2 Účelová funkce

Účelová funkce je definována jako průměr přesností na validační množině. Pro každé řešení je natrénováno sedm neuronových sítí, u kterých se dynamicky mění počet neuronů ve skryté vrstvě. Podle pravidla (rovnice 6.1) „Geometric Pyramid Rules“ se stanoví počáteční počet neuronů ve skryté vrstvě  $H$ .

$$H = N * M \quad (6.1)$$

Kde  $N$  je počet vstupních atributů a  $M$  je počet neuronů ve výstupní vrstvě. Pro každé další trénování se přičítá nebo odečítá předem stanovený počet neuronů. Výsledný průměr přesností se vytvoří ze sítí, které mají ve skryté vrstvě následující počet neuronů ( $H, H+1, H-1, H+2, H-2, H+3, H-3$ ). Přičítání takto pevného počtu neuronů bylo stanovena na základě série experimentů, kde pro větší počet neuronů než  $H+3$  docházelo k přeučení modelu a pro menší počet neuronů než  $H-3$  docházelo zase k podučení neuronové sítě.

## 6.4 Výběr sázkové strategie

V literatuře [1, 5] se vyskytuje několik sázkových strategií. V této kapitole jsou popsány pouze tři základní, které jsou v simulační části práce pro každý model mezi sebou porovnány. Všechny strategie mají společnou vlastnost, že je sázka vsazena jen v případě, kdy se jedná o hodnotný kurz (kurz vypsaný modelem je nižší než kurz vypsaný sázkovou kanceláří).

### 6.4.1 Fixní sázka

Fixní sázka je jednoduchá strategie, při které se každé hodnotné sázce přiřadí 1 % - 5 % bankrollu (veškeré peníze, které má sázkař vyhrazené na sázení) [1].

### 6.4.2 Fixní návratnost

Fixní návratnost je strategie, při které je pevně zvolen zisk z každé vsazené sázky [1].

Výše sázky na vybranou hodnotnou sázku je popsána vztahem (6.2).

$$f = \frac{P}{b - 1} \quad (6.2)$$

Kde  $f$  znamená, kolik prostředků je zapotřebí vsadit na vybranou hodnotnou sázku, aby bylo dosaženo stanoveného zisku  $P$  a  $b$  je kurz sázkové kanceláře v desetinné podobě.

### 6.4.3 Kellyho kritérium

Kellyho kritérium [24] je vzorec (6.3) pro určení optimální velikosti sázky.

$$f^* = \frac{bp - q}{b - 1} \quad (6.3)$$

Kde  $f^*$  znamená vyjádření, kolik procent z bankrollu je třeba vsadit na hodnotnou sázku,  $b$  je kurz sázkové kanceláře v desetinné podobě,  $p$  je pravděpodobnost úspěchu (desetinný zápis) vyčíslená modelem a  $q$  je pravděpodobnost neúspěchu ( $1 - p$ ).

Problémem Kellyho kritéria je, že pokud se pravděpodobnost stanovená modelem od pravděpodobnosti bookmakera velmi liší, mohl by sázkař vsadit i více jak 50 % bankrollu, a proto se doporučuje sázet jen pevně stanovenou část Kellyho kritéria. Pro modelový příklad, který je uveden v tabulce (Tabulka 2) byl zvolen kurz bookmakera 1,8 (odpovídá pravděpodobnosti 0,556) na výhru domácího týmu a pravděpodobnost výhry domácího týmu z modelu neuronové sítě 0,625 (odpovídá kurzu 1,6)

Tabulka 2 – Porovnání Kellyho kritéria

Typ kritéria	$f^*[\%]$
Plné kritérium	15,63
1/2 kritéria	7,815
1/4 kritéria	3,908
1/10 kritéria	1,563

Z tabulky (Tabulka 2) je možné pozorovat, že při rozdílu 6,9 % by bylo nutné vsadit 15,63 % bankrollu, a proto byla zvolena 1/10 Kellyho kritéria pro simulační část práce, která snižuje míru rizika bankrotu.

## 6.5 Sázkový simulátor

Pro otestování ziskovosti predikčních modelů pomocí 3 sázkových strategií, které jsou popsány v předchozí kapitole, byl vytvořen v programovacím jazyku Python script pro simulaci sázení. Simulátor sázení simuluje sezónu 2020/21 anglické fotbalové nejvyšší soutěže Premier League. Zápasy jsou nabízeny modelu ke vsazení v pořadí, jak byly odehrány v reálném životě. Pro každý zápas jsou k dispozici historické kurzy sazkové kanceláře Tipsport. Při reálném použití by bylo ještě třeba brát v potaz kurzy konkurenčních sázkových kanceláří a v případě hodnotného kurzu si vsadit u kanceláře, která nabízí nejvyšší kurz na vybranou příležitost a tím maximalizovat zisk. Každá simulace začíná s počátečním bankrollem 100 jednotek a simuluje zároveň všechny 3 sázkové strategie.

## 7 EXPERIMENTY A VÝSLEDKY

V této části práce je popsáno nastavení jednotlivých modelů, které byly použity pro sázení hodnotných kurzů na sázku *více rohů v zápasu* o poločasové přestávce. U jednotlivých modelů je každá sázková strategie porovnána z hlediska návratnosti investice (ROI) pro předem stanovené omezení. Horní hranice se v průběhu simulací mění a znázorňuje maximální pravděpodobnost, kdy si model vsadí, pokud se jedná o hodnotný kurz. Spodní hranice je nastavena na 1,2 a udává minimální kurz u sázkové kanceláře. Pokud je hodnotný kurz nižší než tato hranice, model tuto hodnotnou sázku nevsadí. Ke každé sázkové strategii jsou vykresleny grafy pro jednotlivá omezení, které zobrazují průběh ROI během sezony 2020/2021 anglické Premier League.

Pro vyhodnocení modelu neuronové sítě se používají 3 sázkařské strategie:

- Fixní sázka, která je nastavena na 2 % bankrollu
- Fixní návratnost, kde fixní zisk je nastaven na 2 jednotky
- 1/10 Kellyho kritéria

Každé sázkové strategii byl nastaven počáteční bankroll na 100 jednotek.

Pro každý model je vytvořen speciální dataset, ze kterého se pomocí Genetického algoritmu vybírá pouze podmnožina atributů.

Pro oba datasey čítá trénovací množina 1300 vzorků odehraných zápasů z nejvyšší anglické fotbalové soutěže ze sezon 2015/16, 2016/2017 a 2017/2018. Zápasy, ve kterých jeden z týmů vedl o poločasové přestávce o více než 4 rohové kopy nebo jeden z týmů dostal v prvním poločase červenou kartu, byly z datasetu odstraněny. Validační sadu proto tvoří pouze 322 vzorků ze sezony 2019/2020.

Testovací dataset je tvořen 141 zápasy odehranými v období od 16. 12. 2020 do 10. 5. 2021, u kterých byly přímo sbírány sázkové kurzy u společnosti Tipsport. Bylo proto možné vyhodnotit i reálnou úspěšnost finanční bilance a také porovnat jednotlivé navržené modely s modelem společnosti Tipsport z hlediska přesnosti.

## 7.1 Model I

Pro model I bylo vybráno pomocí Genetického algoritmu 58 atributů z celkového počtu 129 vytvořených atributů. Při výběru podmnožiny atributů se autor práce setkal s problémem, že byl atribut vybrán pouze pro domácí (hostující) tým. Tento problém byl odstraněn pomocí manuálního přidání chybějícího atributu i pro tým soupeře. 42 atributů je z kategorie poločasové statistiky a zbylé 2 jsou z kategorie předzápasových kurzů. Vybrané atributy jsou detailně popsány v příloze PI a vytvořený dataset pro model I je také dostupný na příloženém CD.

### 7.1.1 Použité nastavení

*Tabulka 3 – Použité parametry Modelu I*

Počet neuronů ve vrstvách	44–14–3
Aktivační funkce (skrytá vrstva)	ReLU
Aktivační funkce (výstupní vrstva)	Softmax
L2 regularizace	0,1
Velikost dávky	128
Optimalizátor	Adam
Ztrátová funkce	Kategorická křížová entropie
Normalizace dat	Normalizace směrodatnou odchylkou

Jelikož se významně lišilo rozmezí jednotlivých atributů byla provedena normalizace dat pomocí směrodatné odchylky. Je důležité také zmínit, že hodnoty pro normalizaci validační a testovací množiny jsou vypočítány pouze z množiny trénovací.

Byla použita neuronová síť s jednou skrytou vrstvou, která měla ve vstupní vrstvě 44 neuronů, které značí stejný počet vstupních atributů. Pro skrytou vrstvu bylo nastaveno 14 neuronů s aktivační funkcí ReLU a váhovou L2 regularizací 0,1. Výstupní vrstva se skládala ze 3 neuronů a aktivační funkce Softmax, která vrací pravděpodobnost pro všechny 3 možné výsledky, tedy výhru domácího týmu, hostujícího týmu a remízu. Jako ztrátová funkce je použita kategorická křížová entropie. Dalším hyperparametrem je velikost dávky, která byla nastavena na hodnotu 128 a znamená to, že neuronovou sítí projde 128 tréninkových



příkladů, než dojde k úpravě vah, takže jsou váhy upravovány několikrát během jedné epochy.

### 7.1.2 Výsledky sázkových strategií

Tabulka 4 – Porovnání pro sázkovou strategii fixní sázka

Omezení	Průměrné $ROI_{\text{Fixní\_sázka}}$	Minimální $ROI_{\text{Fixní\_sázka}}$	Maximální $ROI_{\text{Fixní\_sázka}}$	Průměrný počet hodnotných sázek	Průměrná přesnost hodnotných sázek [%]
$x \leq 3$	0,602	0,57	0,63	34,4	61,53
$x \leq 2$	0,476	0,36	0,69	26,2	76,61
$x \leq 1,8$	0,424	0,34	0,46	22,6	87,75

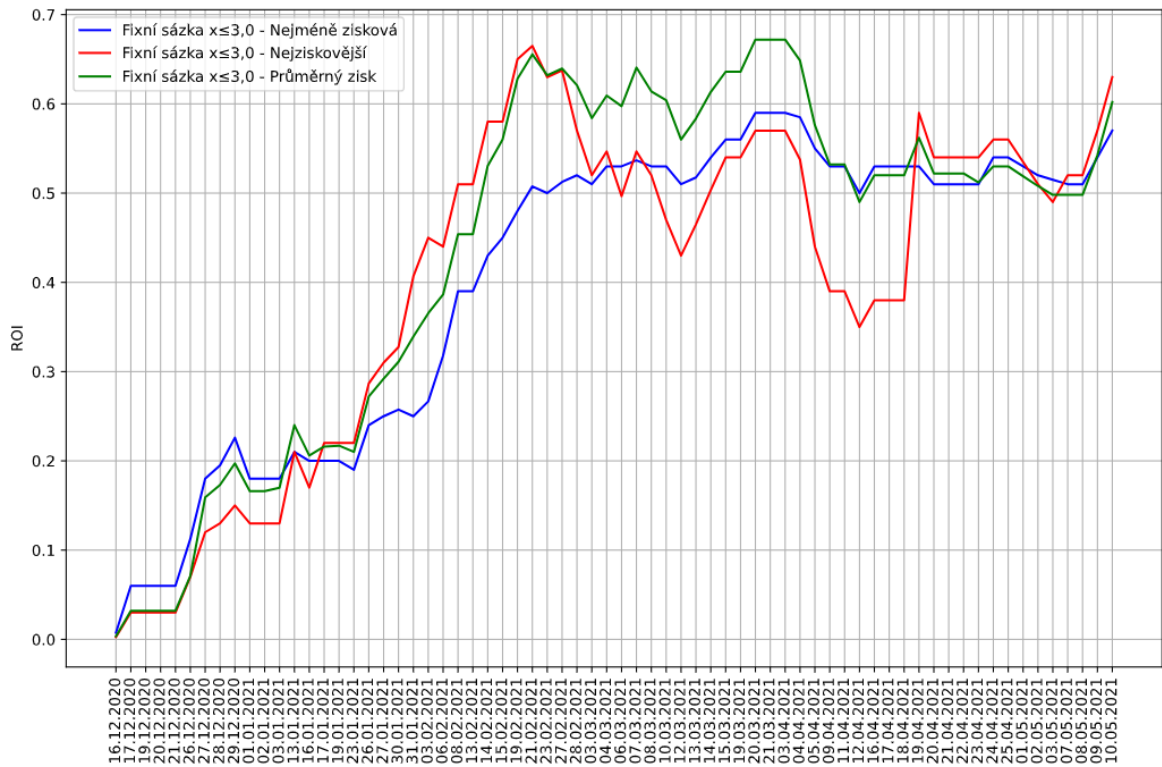
V tabulce (Tabulka 4) je zobrazen souhrn návratnosti investic (ROI) pro sázkovou strategii fixní sázka. Tabulka zobrazuje pro každé omezení konečnou hodnotu ROI pro nejméně ziskovou a nejvíce ziskovou simulaci spolu s průměrnou konečnou návratností investic. Pro všechna zobrazení je také zaznamenán průměrný počet vsazených sázek a jejich přesnost.

Z tabulky lze pozorovat trend, že při snižování hranice omezení dochází ke vsazení menšího počtu sázek, ale dochází ke zvýšení přesnosti. Maximální hodnota ROI na konci sezony byla dosažena pro omezení  $x \leq 2$ . Nejvyšší průměrná hodnota ROI je pro omezení  $x \leq 3$ . Nejnížší hodnota ROI na konci sezony je u hranice  $x \leq 1,8$ . Největší rozdíl mezi nejnížší a nejvyšší hodnotou ROI je vidět pro ohraničení  $x \leq 2$ , což je dáno tím, že u tohoto omezení je hodně zápasů na hranici a mezi vsazením a nevsazením rozhoduje pouze pár desetin ve vypsaném kurzu neuronovou sítí.

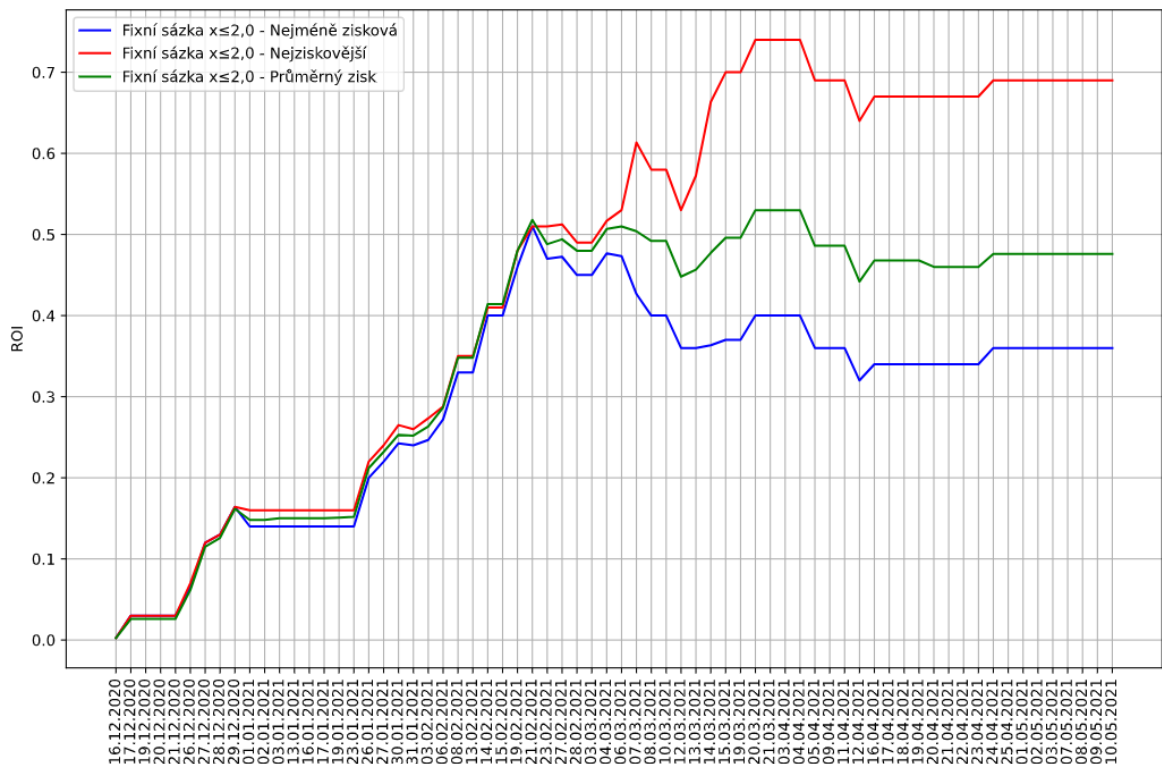
Na obrázku (Obrázek 11) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii fixní sázky a omezení  $x \leq 3$ .

Na obrázku (Obrázek 12) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii fixní sázky a omezení  $x \leq 2$ .

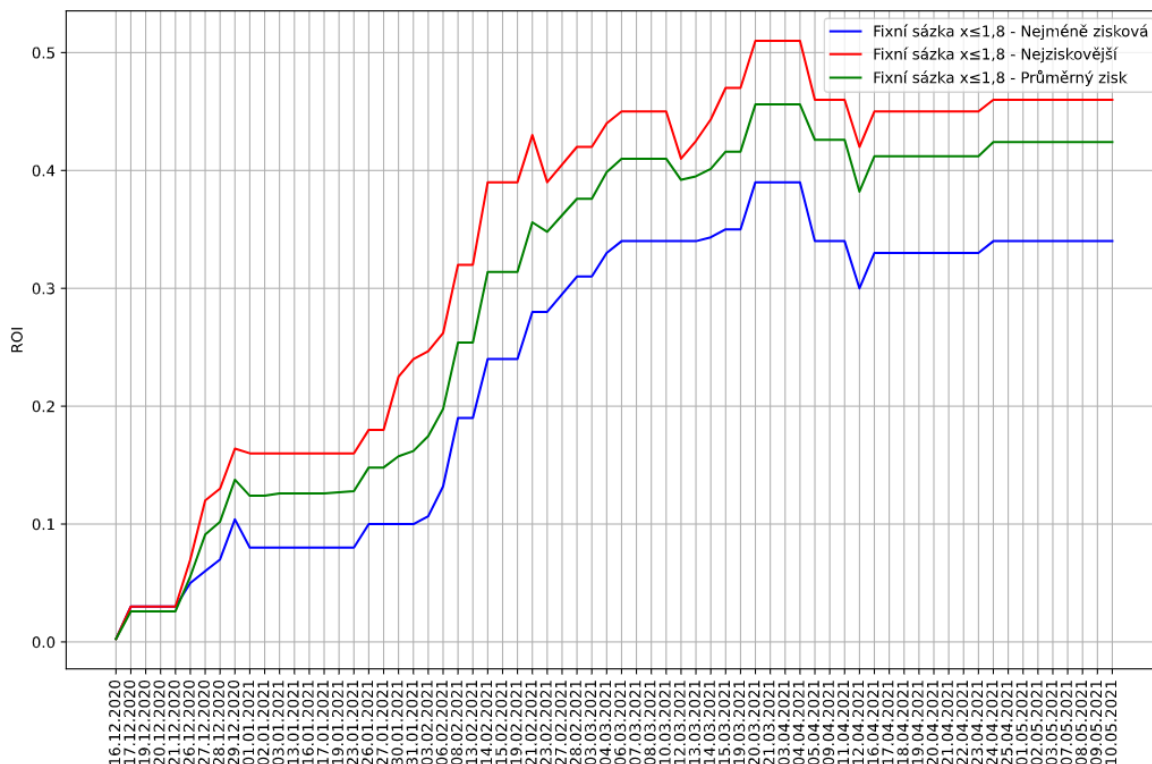
Na obrázku (Obrázek 13) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii fixní sázky a omezení  $x \leq 1,8$ .



Obrázek 11 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky ( $x \leq 3$ )



Obrázek 12 - Porovnání průběhů ROI pro sazkovou strategii fixní sázky ( $x \leq 2$ )


 Obrázek 13 - Porovnání průběhů ROI pro sázkovou strategii fixní sázky ( $x \leq 1,8$ )

Tabulka 5 - Porovnání pro sázkovou strategii fixní výhry

Omezení	Průměrné $ROI_{\text{Fixní}_v\acute{y}hra}$	Minimální $ROI_{\text{Fixní}_V\acute{y}hra}$	Maximální $ROI_{\text{Fixní}_V\acute{y}hra}$	Průměrný počet hodnotných sázek	Průměrná přesnost hodnotných sázek [%]
$x \leq 3,0$	0,638	0,57	0,7	34,4	61,53
$x \leq 2,0$	0,574	0,5	0,64	26,2	76,61
$x \leq 1,8$	0,424	0,34	0,46	22,6	87,75

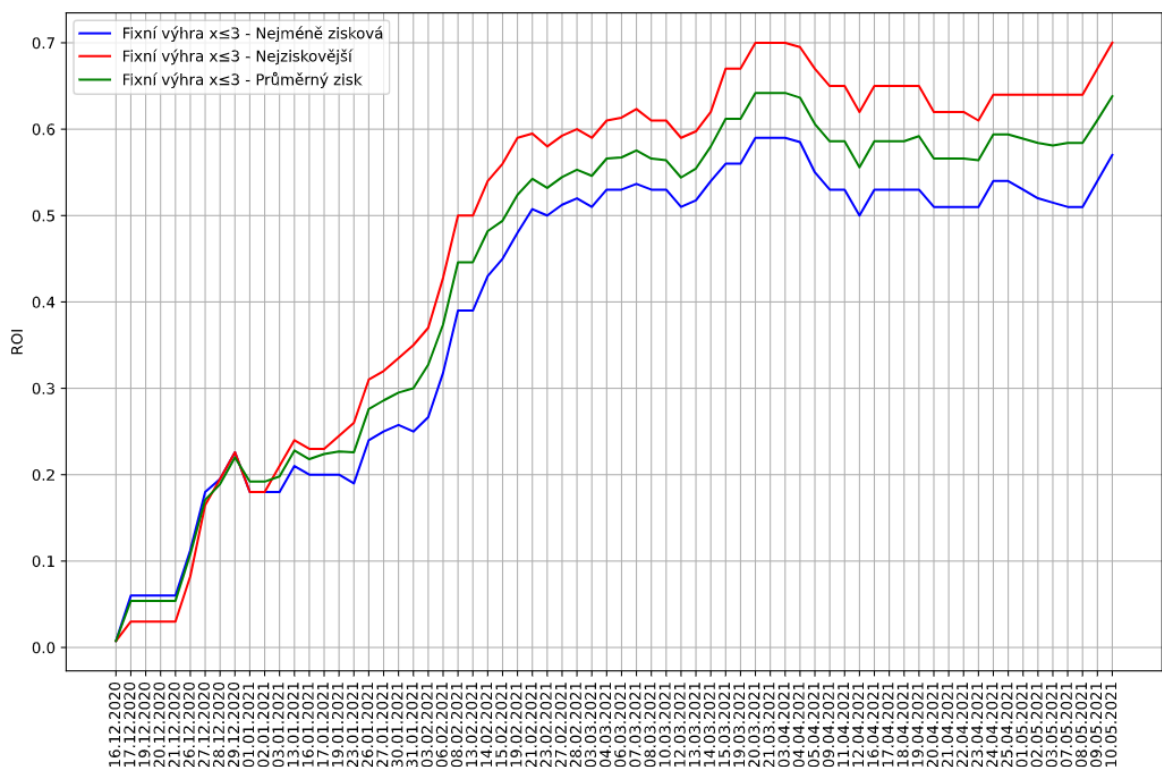
V tabulce (Tabulka 5) je zobrazen souhrn návratnosti investic (ROI) pro sázkovou strategii fixní výhry. Tabulka zobrazuje pro každé omezení konečnou hodnotu ROI pro nejméně ziskovou a nejvíce ziskovou simulaci spolu s průměrnou konečnou návratností investic. Pro všechna zobrazení je také zaznamenán průměrný počet vsazených sázek a jejich přesnost. Z tabulky lze pozorovat trend, že při snižování hranice omezení dochází ke vsazení menšího počtu sázek, ale dochází ke zvýšení přesnosti. Maximální hodnota ROI, ale i průměrná

hodnota ROI na konci sezony byla dosažena pro omezení  $x \leq 3$ . Nejnižší hodnota ROI na konci sezony je u hranice  $x \leq 1,8$ . Největší rozdíl mezi nejnižší a nejvyšší hodnotou ROI je vidět pro ohraničení  $x \leq 2$ , což je dáno tím, že u tohoto omezení je hodně zápasů na hranici a mezi vsazením a nevsazením rozhoduje pouze pár desetín ve vypsaném kurzu neuronovou sítí.

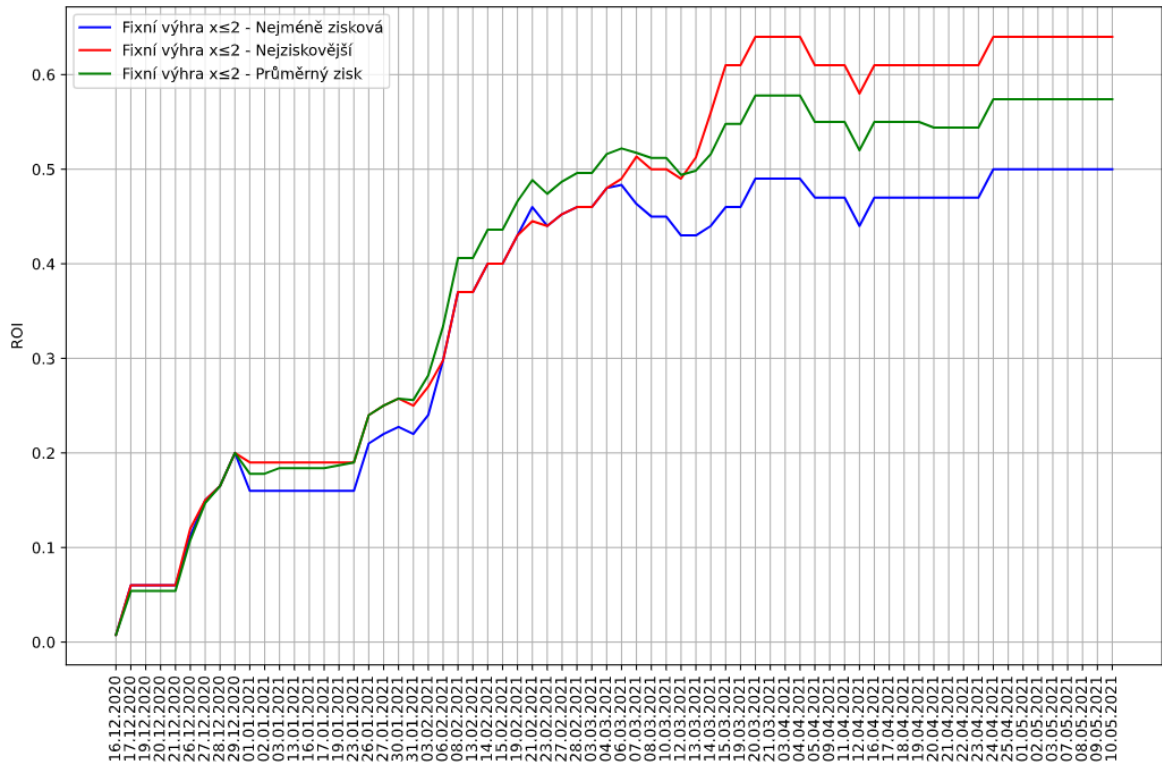
Na obrázku (Obrázek 14) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii fixní výhry a omezení  $x \leq 3$ .

Na obrázku (Obrázek 15) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii fixní výhry a omezení  $x \leq 2$ .

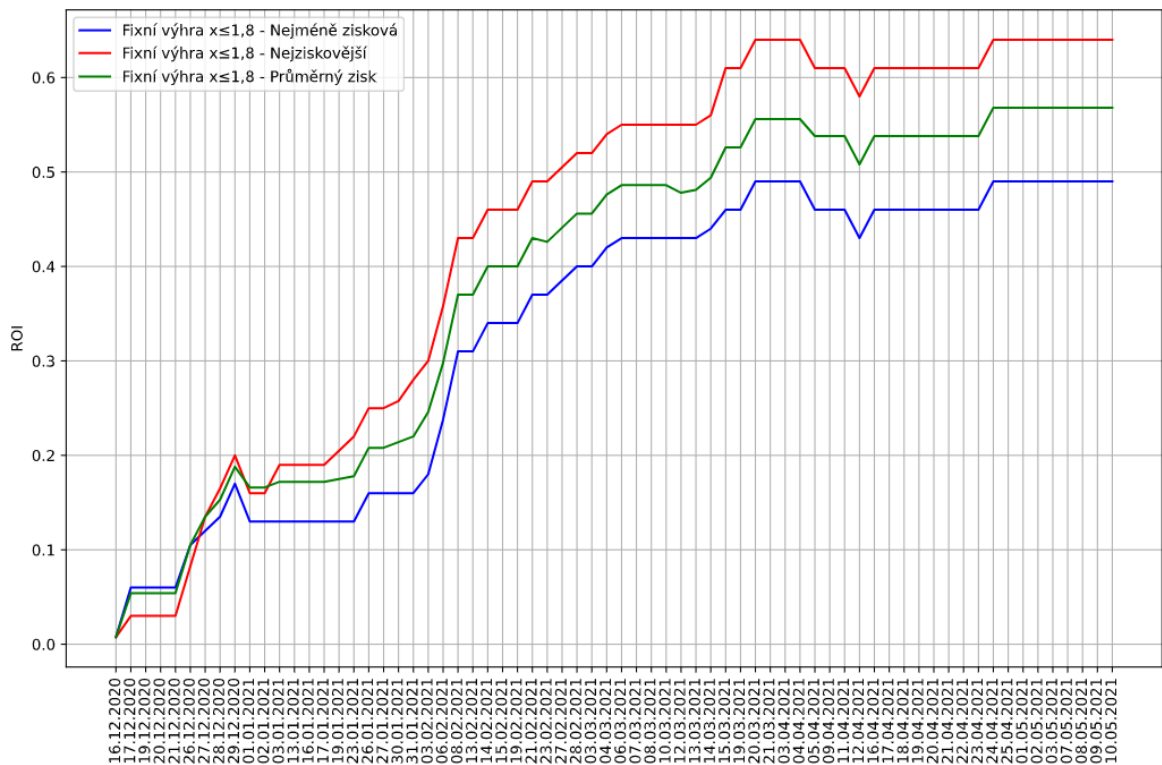
Na obrázku (Obrázek 16) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii fixní výhry a omezení  $x \leq 1,8$ .



Obrázek 14 - Porovnání průběhů ROI pro sázkovou strategii fixní výhry ( $x \leq 3$ )



Obrázek 15 - Porovnání průběhů ROI pro sazkovou strategii fixní sázky ( $x \leq 2$ )



Obrázek 16 - Porovnání průběhů ROI pro sazkovou strategii fixní sázky ( $x \leq 1,8$ )

Tabulka 6 - Porovnání pro sázkovou strategii 10 % Kellyho kritéria

Omezení	Průměrné ROI <sub>1/10Kelly</sub> yho_kritéria	Minimální ROI <sub>1/10Kelly</sub> ho_kritéria	Maximální ROI <sub>1/10Kellyho_</sub> kritéria	Průměrný počet hodnotných sázek	Průměrná přesnost hodnotných sázek [%]
$x \leq 3$	0,462	0,27	0,71	34,4	61,53
$x \leq 2$	0,4	0,28	0,67	26,2	76,61
$x \leq 1,8$	0,326	0,26	0,41	22,6	87,75

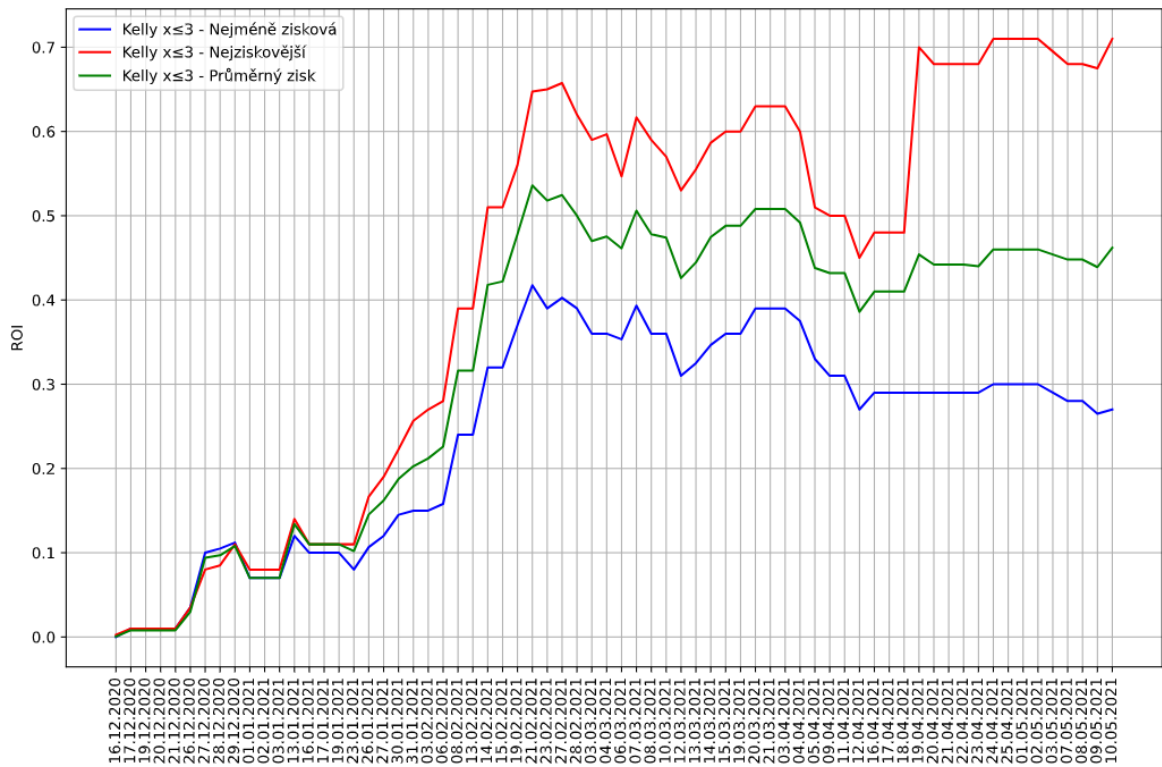
V tabulce (Tabulka 6) je zobrazen souhrn návratnosti investic (ROI) pro sázkovou strategii 10 % Kellyho kritéria. Tabulka zobrazuje pro každé omezení konečnou hodnotu ROI pro nejméně ziskovou a nejvíce ziskovou simulaci spolu s průměrnou konečnou návratností investic. Pro všechna zobrazení je také zaznamenán průměrný počet vsazených sázek a jejich přesnost.

Z tabulky lze pozorovat trend, že při snižování hranice omezení dochází ke vsazení menšího počtu sázek, ale dochází ke zvýšení přesnosti. Maximální hodnota ROI na konci sezony byla dosažena pro omezení  $x \leq 3$ . Nejvyšší průměrná hodnota ROI je pro omezení  $x \leq 3$ . Nejnižší hodnota ROI na konci sezony je u hranice  $x \leq 1,8$ . Největší rozdíl mezi nejnižší a nejvyšší hodnotou ROI je vidět pro ohraničení  $x \leq 3$ , ale i pro omezení  $x \leq 2$  je rozdíl více než 35 %, což je dáno tím, že u této sázkové strategie se dynamicky mění vsazený počet jednotek na základě rozdílů kurzů mezi bookmakerem a neuronovou sítí.

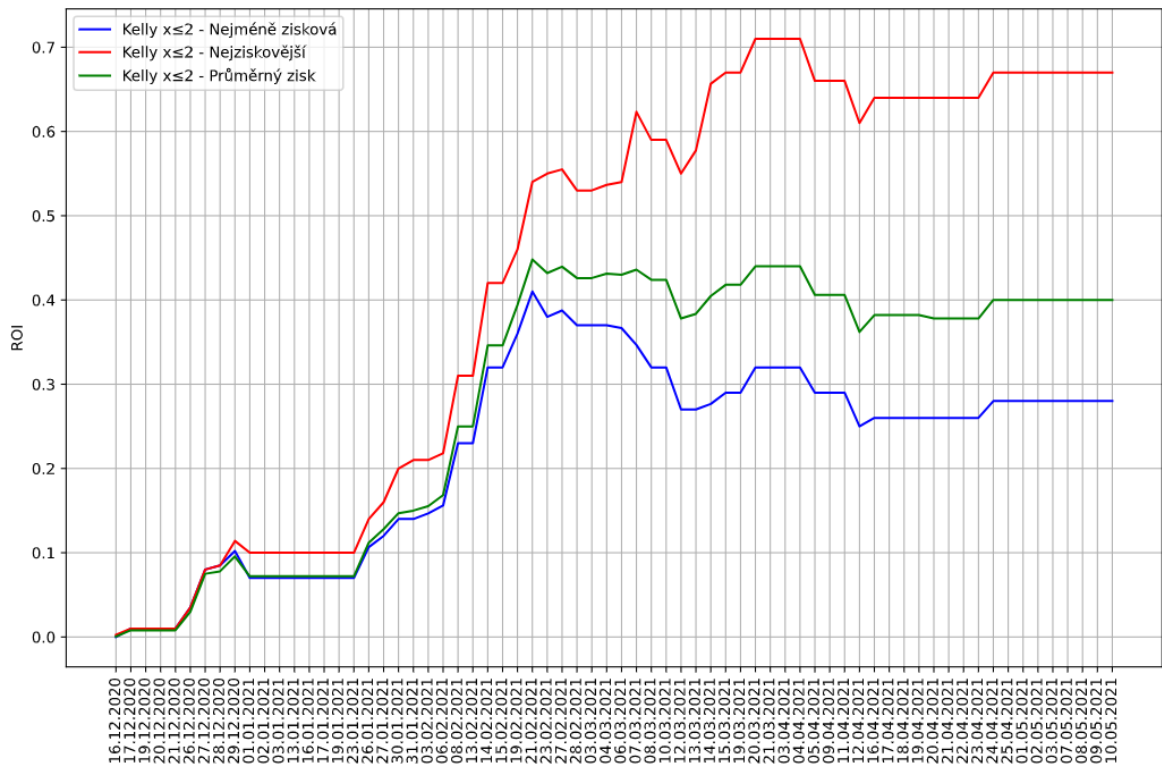
Na obrázku (Obrázek 17) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii 10 % Kellyho kritéria a omezení  $x \leq 3$ .

Na obrázku (Obrázek 18) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii 10 % Kellyho kritéria a omezení  $x \leq 2$ .

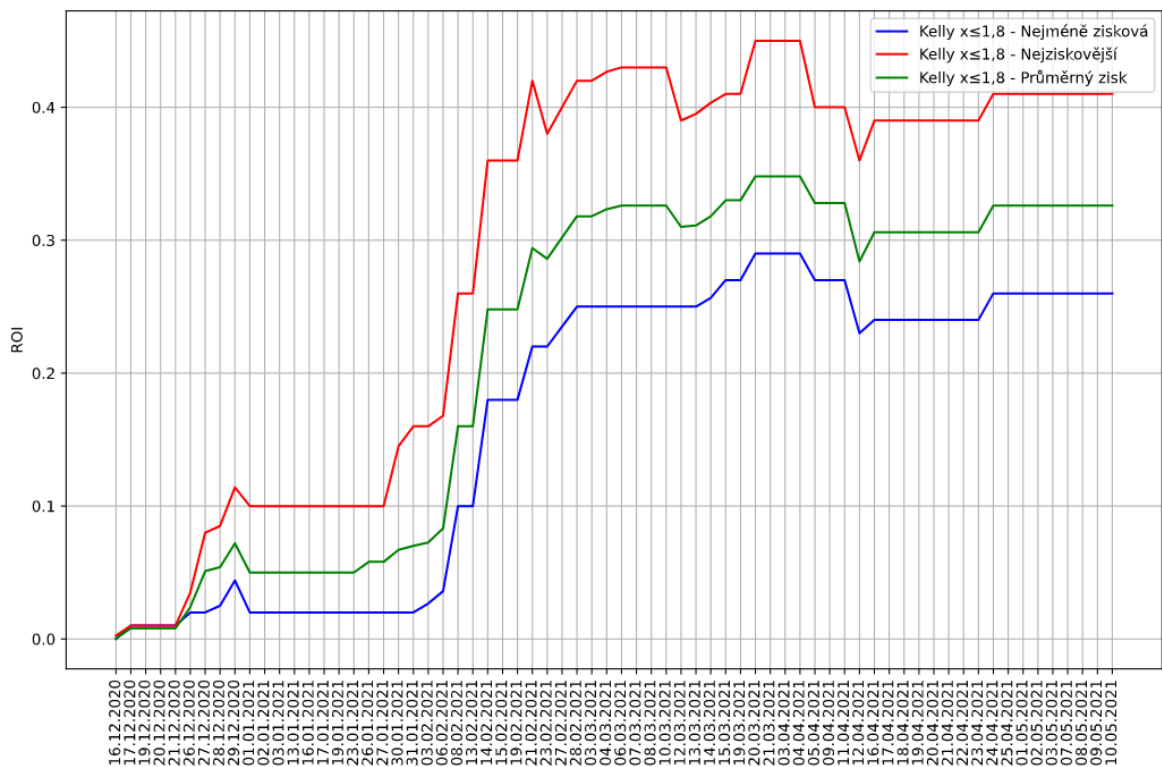
Na obrázku (Obrázek 19) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii 10 % Kellyho kritéria a omezení  $x \leq 1,8$ .



Obrázek 17- Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria ( $x \leq 3$ )



Obrázek 18 - Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria ( $x \leq 2$ )



Obrázek 19 - Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria ( $x \leq 1,8$ )

## 7.2 Model II

Model II by se dal taky nazvat modelem rozdílovým, jelikož pro každou dvojici atributů je na základě odečtení atributu hostujícího týmu od atributu domácího týmu vytvořen nový atribut. Například nový atribut *počet rohových kopů* vznikne jako *počet rohových kopů domácího týmu - (mínus) počet rohových kopů hostujícího týmu*. Stejná analogie je uplatněna pro všechny atributy z kategorie poločasová statistika. Kategorie předzápasových kurzů zůstává nezměněna. Pro model II bylo vybráno pomocí Genetického algoritmu 24 atributů z celkového počtu 63. Vybrané atributy jsou detailně popsány v příloze PII a vytvořený dataset (podmnožina atributů) pro model II je také dostupný na příloženém CD.

### 7.2.1 Použité nastavení

Tabulka 7 – Použité parametry Modelu II

Počet neuronů ve vrstvách	24–10–3
Aktivační funkce (skrytá vrstva)	ReLU
Aktivační funkce (výstupní vrstva)	Softmax



L2 regularizace	0,1
Velikost dávky	32
Optimalizátor	Adam
Ztrátová funkce	Binární křížová entropie
Normalizace dat	Normalizace směrodatnou odchylkou

Jelikož se významně lišilo rozmezí jednotlivých atributů byla provedena normalizace dat pomocí směrodatné odchylky. Je důležité také zmínit, že hodnoty pro normalizaci validační a testovací množiny jsou vypočítány pouze z množiny trénovací.

Byla použita neuronová síť s jednou skrytou vrstvou, která měla ve vstupní vrstvě 24 neuronů, které značí stejný počet vstupních atributů. Pro skrytou vrstvu bylo nastaveno 10 neuronů s aktivační funkcí ReLU a váhovou L2 regularizací 0,1. Výstupní vrstva se skládala ze 3 neuronů a aktivační funkce Softmax, která vrací pravděpodobnost pro všechny 3 možné výsledky, tedy výhru domácího týmu, hostujícího týmu a remízu. Jako ztrátová funkce je použita kategorická křížová entropie. Dalším hyperparametrem je velikost dávky, která byla nastavena na hodnotu 32 a znamená to, že neuronovou sítí projde 32 tréninkových příkladů, než dojde k úpravě vah, takže jsou váhy upravovány několikrát během jedné epochy.

## 7.2.2 Výsledky sázkových strategií

Tabulka 8 - Porovnání pro sázkovou strategii fixní sázka

Omezení	Průměrné $ROI_{\text{Fixní sázka}}$	Minimální $ROI_{\text{Fixní sázka}}$	Maximální $ROI_{\text{Fixní sázka}}$	Průměrný počet hodnotných sázek	Průměrná přesnost hodnotných sázek [%]
$x \leq 3,0$	0,488	0,37	0,58	53	49,49
$x \leq 2,0$	0,252	0,15	0,37	21,4	73,07
$x \leq 1,8$	0,192	0,16	0,25	18,6	74

V tabulce (Tabulka 8) je zobrazen souhrn návratnosti investic (ROI) pro sázkovou strategii fixní sázka. Tabulka zobrazuje pro každé omezení konečnou hodnotu ROI pro nejméně

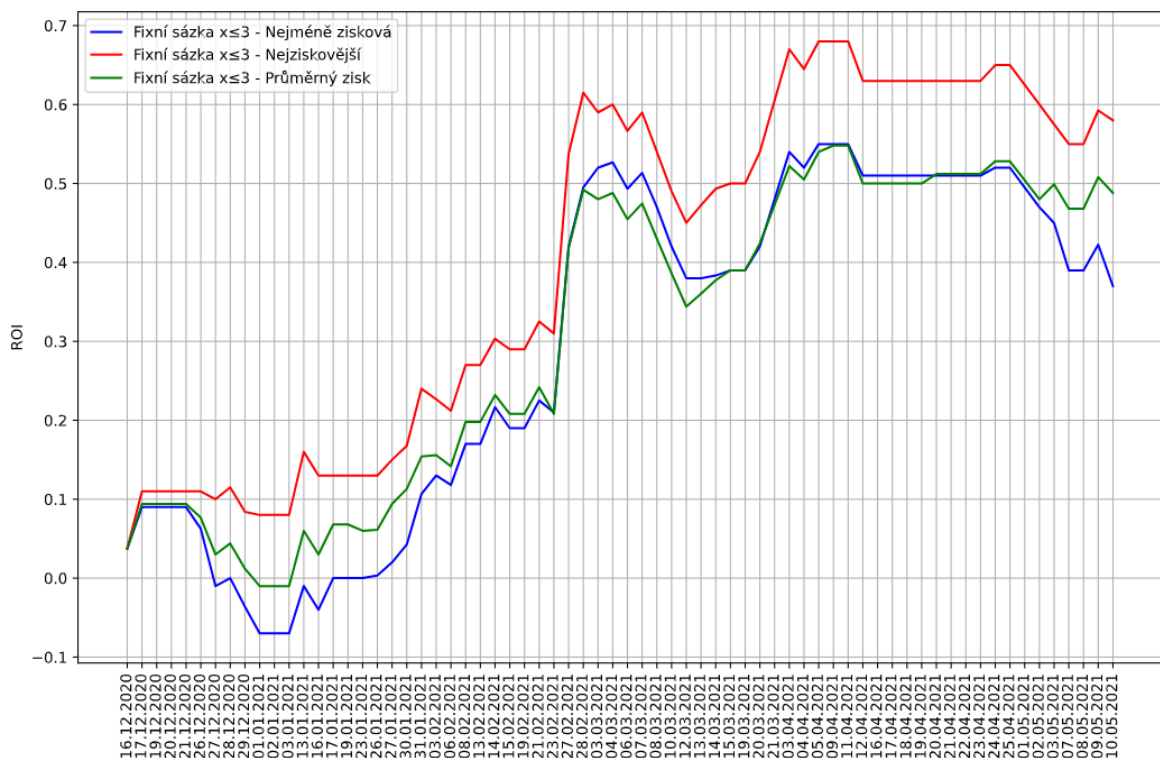
ziskovou a nejvíce ziskovou simulaci spolu s průměrnou konečnou návratností investic. Pro všechna zobrazení je také zaznamenán průměrný počet vsazených sázek a jejich přesnost.

Z tabulky lze pozorovat trend, že při snižování hranice omezení dochází ke vsazení menšího počtu sázek, ale dochází ke zvýšení přesnosti. Maximální hodnota ROI na konci sezony byla dosažena pro omezení  $x \leq 2$ . Nejvyšší průměrná hodnota ROI je pro omezení  $x \leq 3$ . Nejnižší hodnota ROI na konci sezony je u hranice  $x \leq 1,8$ . Největší rozdíl mezi nejvyšší a nejnižší hodnotou ROI je vidět pro ohraničení  $x \leq 2$ , což je dáno tím, že u tohoto omezení je hodně zápasů na hranici a mezi vsazením a nevsazením rozhoduje pouze pár desetin ve vypsání kurzu neuronovou sítí.

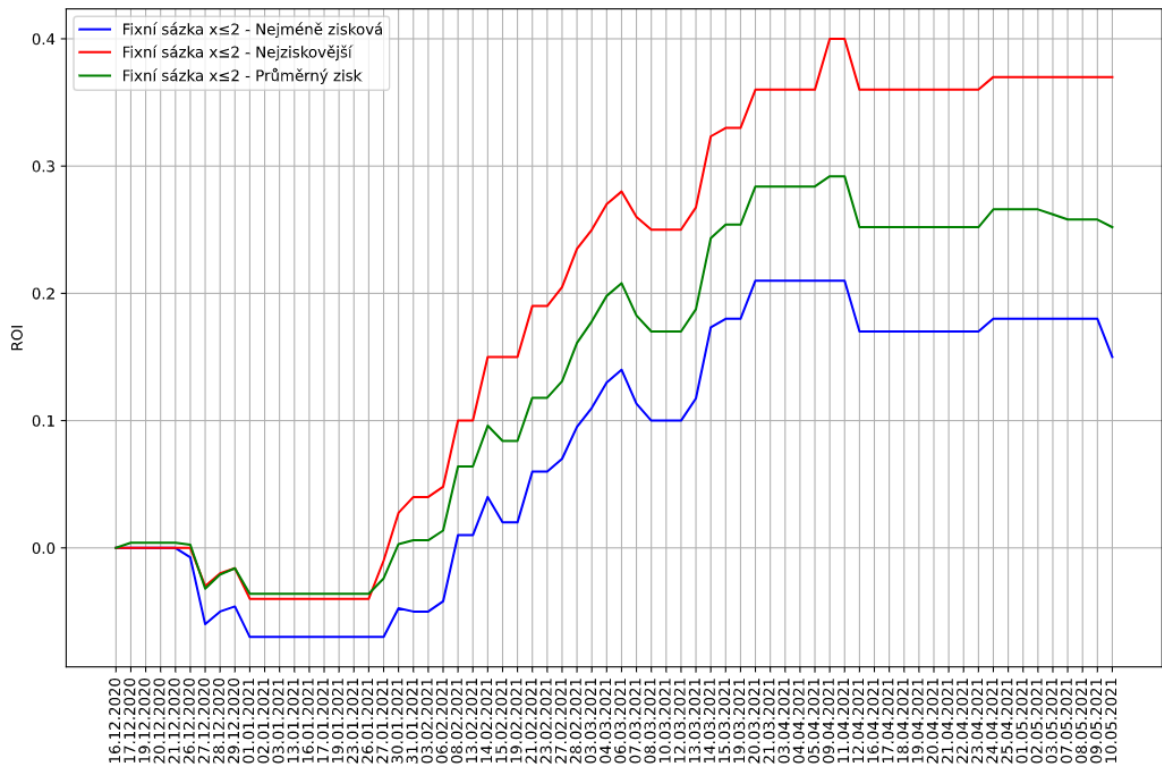
Na obrázku (Obrázek 11) jsou zobrazeny průběhy návratnosti investic v průběhu sezony 2020/2021 anglické Premier League pro sázkovou strategii fixní sázky a omezení  $x \leq 3$ .

Na obrázku (Obrázek 12) jsou zobrazeny průběhy návratnosti investic v průběhu sezony 2020/2021 anglické Premier League pro sázkovou strategii fixní sázky a omezení  $x \leq 2$ .

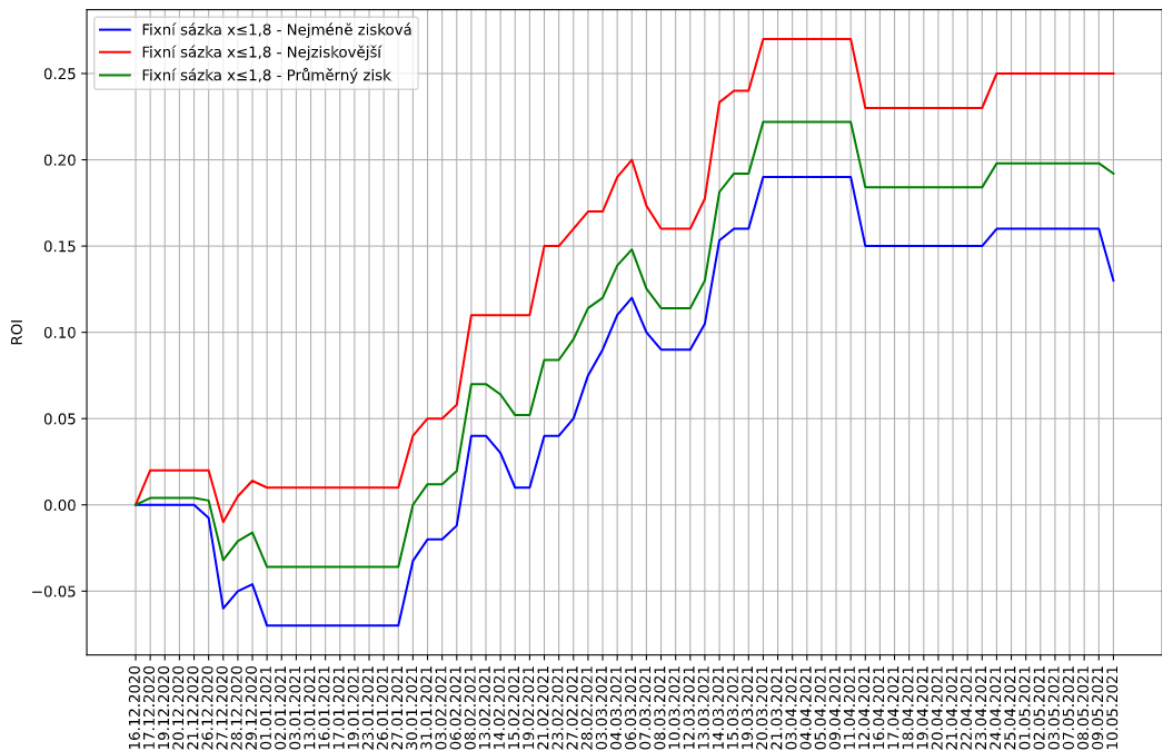
Na obrázku (Obrázek 13) jsou zobrazeny průběhy návratnosti investic v průběhu sezony 2020/2021 anglické Premier League pro sázkovou strategii fixní sázky a omezení  $x \leq 1,8$ .



Obrázek 20 – Porovnání průběhů ROI pro sázkovou strategii fixní sázky ( $x \leq 3$ )



Obrázek 21 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky ( $x \leq 2$ )



Obrázek 22 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky ( $x \leq 1,8$ )

Tabulka 9 - Porovnání pro sázkovou strategii fixní výhra

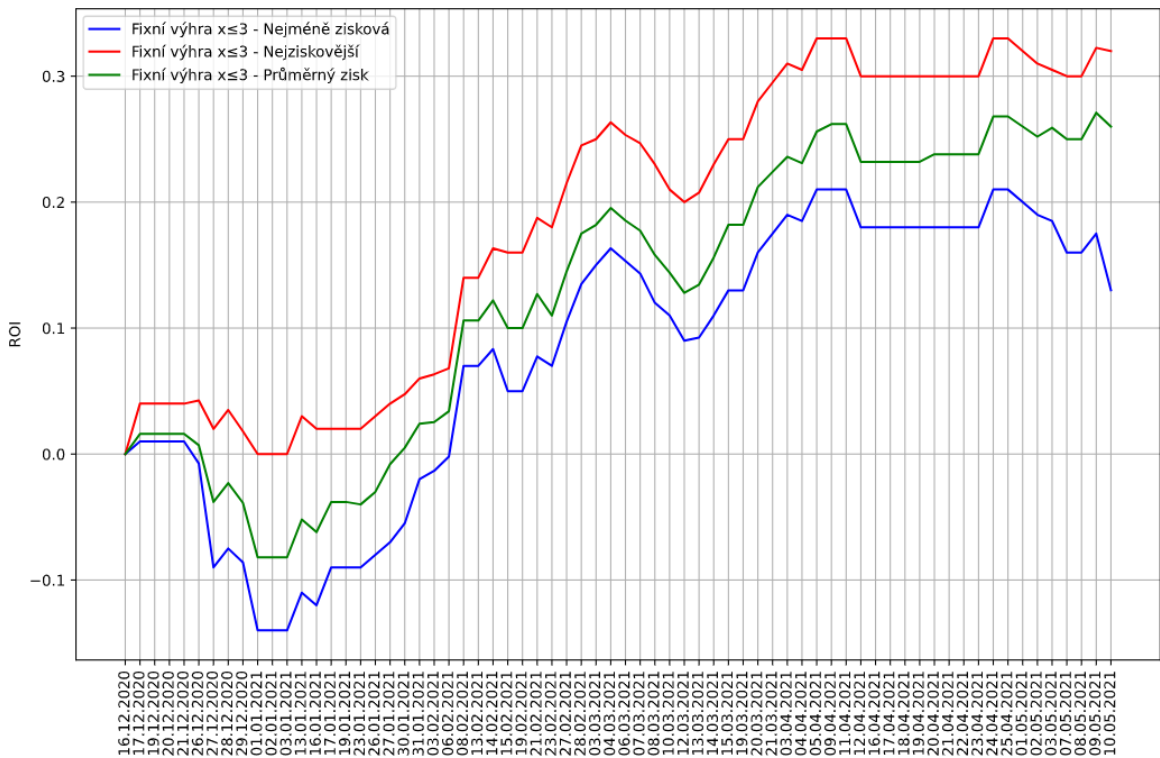
Omezení	Průměrné $ROI_{\text{Fixní}_v\text{ýhra}}$	Minimální $ROI_{\text{Fixní}_v\text{ýhra}}$	Maximální $ROI_{\text{Fixní}_v\text{ýhra}}$	Průměrný počet hodnotných sázek	Průměrná přesnost hodnotných sázek [%]
$x \leq 3,0$	0,26	0,13	0,32	53	49,49
$x \leq 2,0$	0,254	0,14	0,36	21,4	73,07
$x \leq 1,8$	0,212	0,14	0,27	18,6	74

V tabulce (Tabulka 9) je zobrazen souhrn návratnosti investic (ROI) pro sázkovou strategii fixní výhry. Tabulka zobrazuje pro každé omezení konečnou hodnotu ROI pro nejméně ziskovou a nejvíce ziskovou simulaci spolu s průměrnou konečnou návratností investic. Pro všechna zobrazení je také zaznamenán průměrný počet vsazených sázek a jejich přesnost. Z tabulky lze pozorovat trend, že při snižování hranice omezení dochází ke vsazení menšího počtu sázek, ale dochází ke zvýšení přesnosti. Maximální hodnota ROI na konci sezony byla dosažena pro omezení  $x \leq 2$ . Nejvyšší průměrná hodnota ROI je pro omezení  $x \leq 3$ . Největší rozdíl mezi nejnižší a nejvyšší hodnotou ROI je vidět pro ohraničení  $x \leq 2$ , což je dáno tím, že u tohoto omezení je hodně zápasů na hranici a mezi vsazením a nevsazením rozhoduje pouze pár desetin ve vypsání kurzu neuronovou sítí.

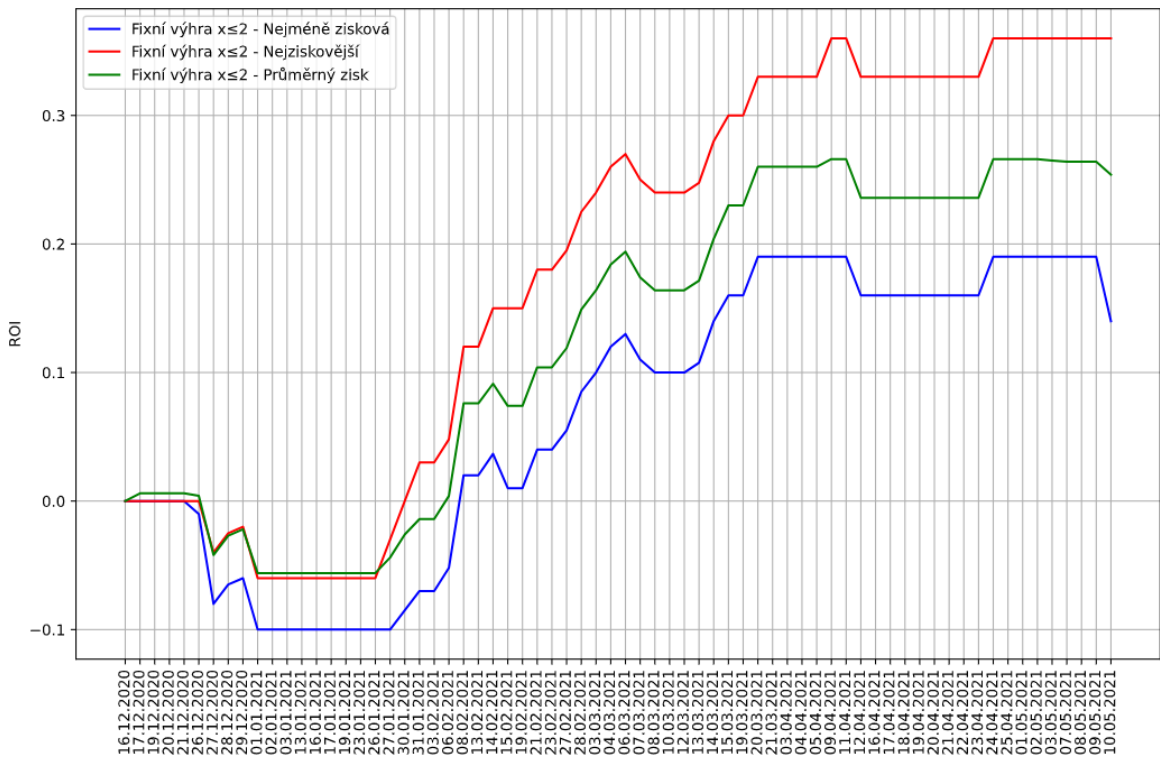
Na obrázku (Obrázek 23) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii fixní výhry a omezení  $x \leq 3$ .

Na obrázku (Obrázek 24) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii fixní výhry a omezení  $x \leq 2$ .

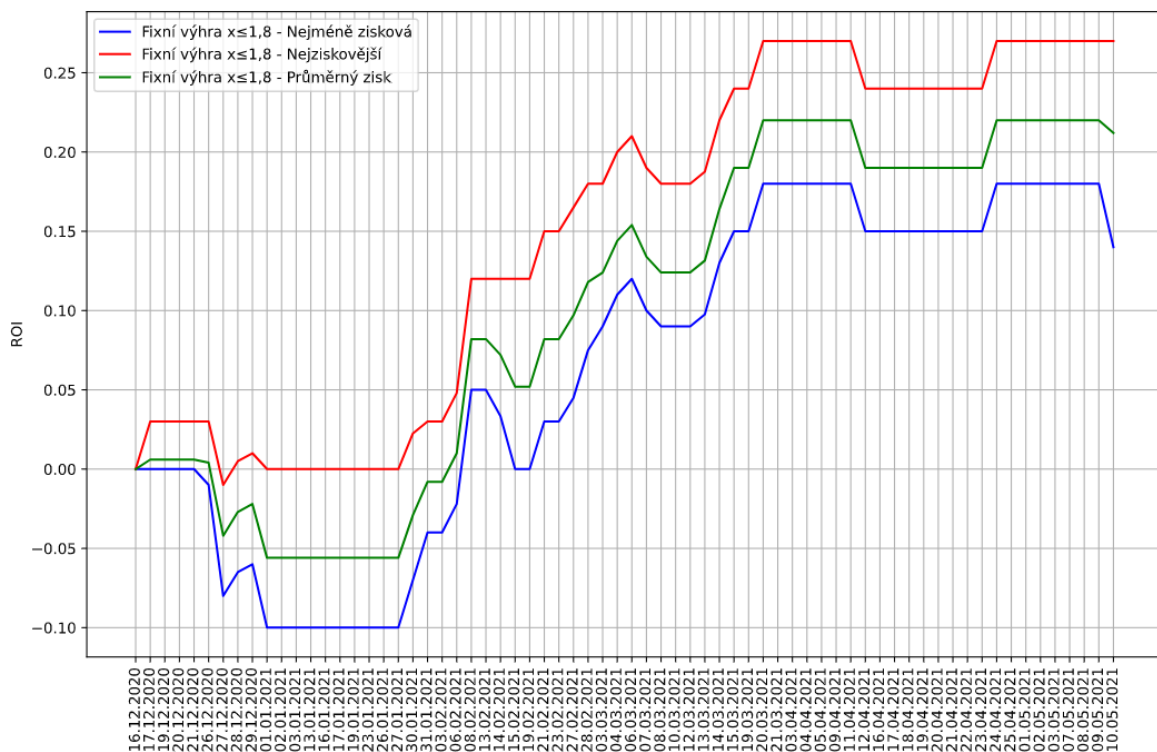
Na obrázku (Obrázek 25) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii fixní výhry a omezení  $x \leq 1,8$ .



Obrázek 23 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky ( $x \leq 3$ )



Obrázek 24 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky ( $x \leq 2$ )


 Obrázek 25 – Porovnání průběhů ROI pro sázkovou strategii fixní sázky ( $x \leq 1,8$ )

Tabulka 10 - Porovnání pro sázkovou strategii 10 % Kellyho kritéria

Omezení	Průměrné ROI <sub>1/10Kellyho_</sub> kritéria	Minimální ROI <sub>1/10Kellyho_</sub> kritéria	Maximální ROI <sub>1/10Kellyho_</sub> kritéria	Průměrný počet hodnotných sázek	Průměrná přesnost hodnotných sázek [%]
$x \leq 3,0$	0,644	0,59	0,71	53	49,49
$x \leq 2,0$	0,298	0,23	0,36	21,4	73,07
$x \leq 1,8$	0,26	0,23	0,27	18,6	74

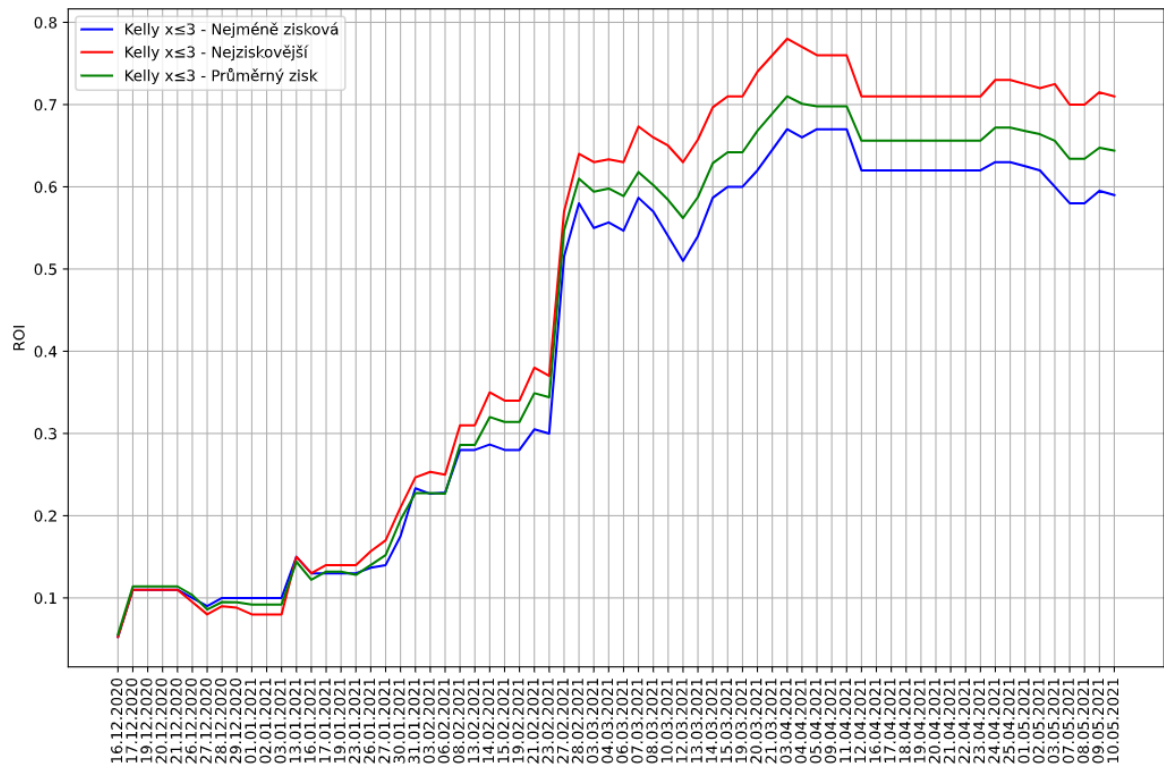
V tabulce (Tabulka 10) je zobrazen souhrn návratnosti investic (ROI) pro sázkovou strategii 10 % Kellyho kritéria. Tabulka zobrazuje pro každé omezení konečnou hodnotu ROI pro nejméně ziskovou a nejvíce ziskovou simulaci spolu s průměrnou konečnou návratností investic. Pro všechna zobrazení je také zaznamenán průměrný počet vsazených sázek a jejich přesnost.

Z tabulky lze pozorovat trend, že při snižování hranice omezení dochází ke vsazení menšího počtu sázek, ale dochází ke zvýšení přesnosti. Maximální hodnota ROI, ale i průměrná hodnota ROI na konci sezony byla dosažena pro omezení  $x \leq 3$ . Nejnížší hodnota ROI na konci sezony je u hranice  $x \leq 1,8$  i  $x \leq 2$ . Největší rozdíl mezi nejnižší a nejvyšší hodnotou ROI je vidět pro ohrazení  $x \leq 2$ , což je dáno tím, že u tohoto omezení je hodně zápasů na hranici a mezi vsazením a nevsazením rozhoduje pouze pár desetin ve vypsáném kurzu neuronovou sítí.

Na obrázku (Obrázek 26) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii 10 % Kellyho kritéria a omezení  $x \leq 3$ .

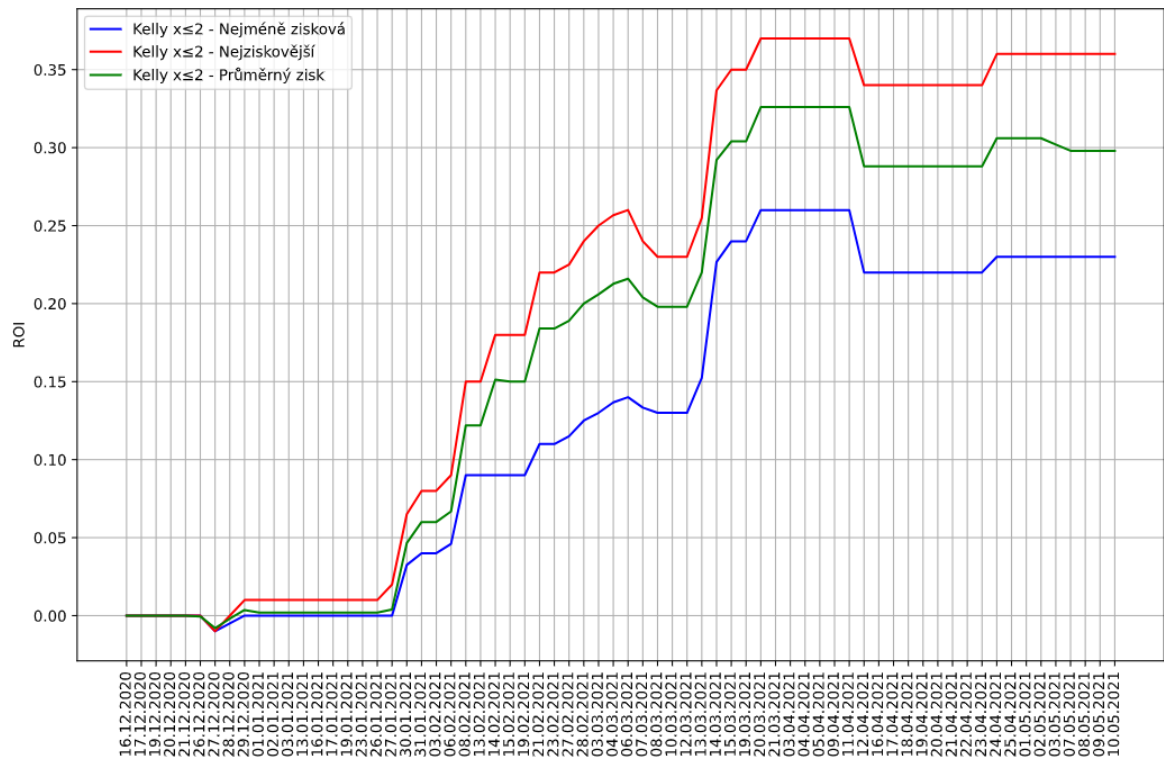
Na obrázku (Obrázek 27) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii 10 % Kellyho kritéria a omezení  $x \leq 2$ .

Na obrázku (Obrázek 28) jsou zobrazeny průběhy návratnosti investic v průběhu sezóny 2020/2021 anglické Premier League pro sázkovou strategii 10 % Kellyho kritéria a omezení  $x \leq 1,8$ .

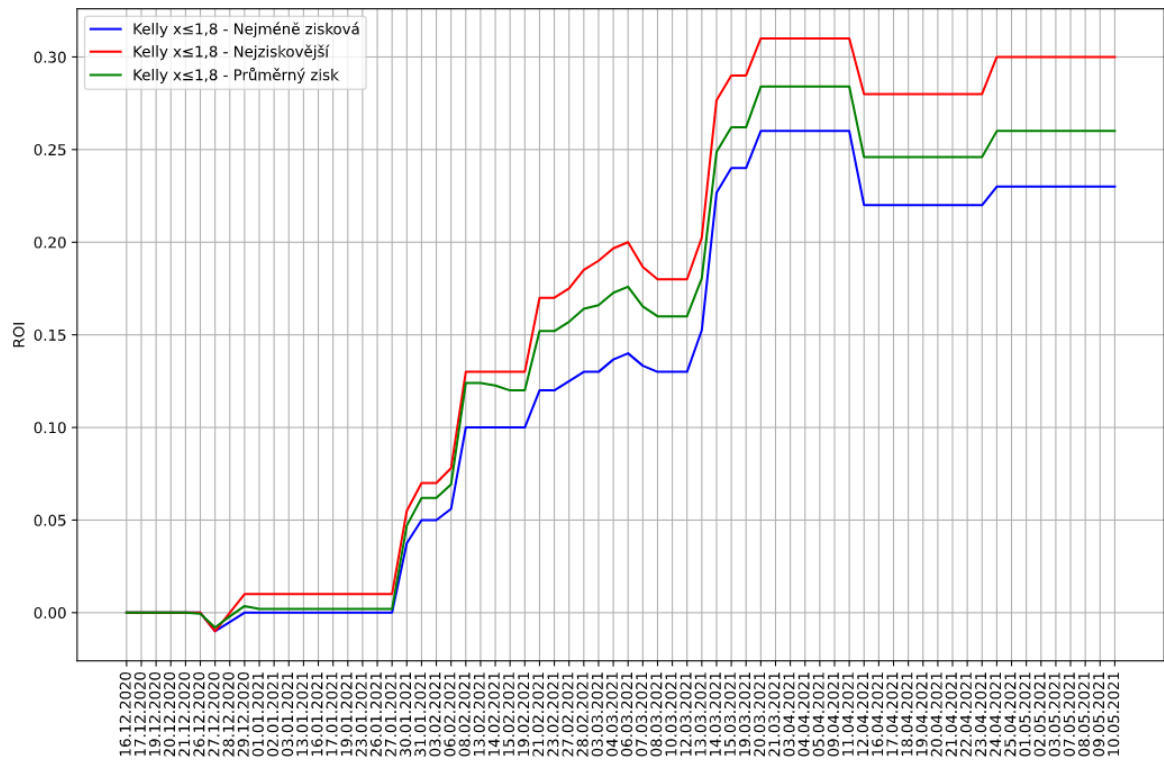


Obrázek 26 - Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria ( $x \leq 3$ )





Obrázek 27 - Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria ( $x \leq 2$ )



Obrázek 28 - Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria ( $x \leq 1,8$ )

## 8 ZHODNOCENÍ VÝSLEDKŮ

V této kapitole se nachází porovnání mezi použitými modely, které byly popsány v předchozí kapitole. Pro každý model byla k porovnání vybrána simulace, která dosahovala nejvyšší přesnosti na validační množině. Pokud byly přesnosti shodné, byla vybrána simulace, která měla nižší ztrátovou funkci. Tento způsob výběru odpovídá reálnému použití na sázkovém trhu. Nejprve se pro každou sázkovou strategii porovnávají oba vytvořené modely z hlediska návratnosti investice (ROI). Poté je pro oba modely vykreslen graf, který pro hodnotné sázky porovnává kurzy společnosti Tipsport s kurzy neuronové sítě.

Nakonec jsou modely porovnány z hlediska přesnosti s modelem sázkové kanceláře Tipsport.

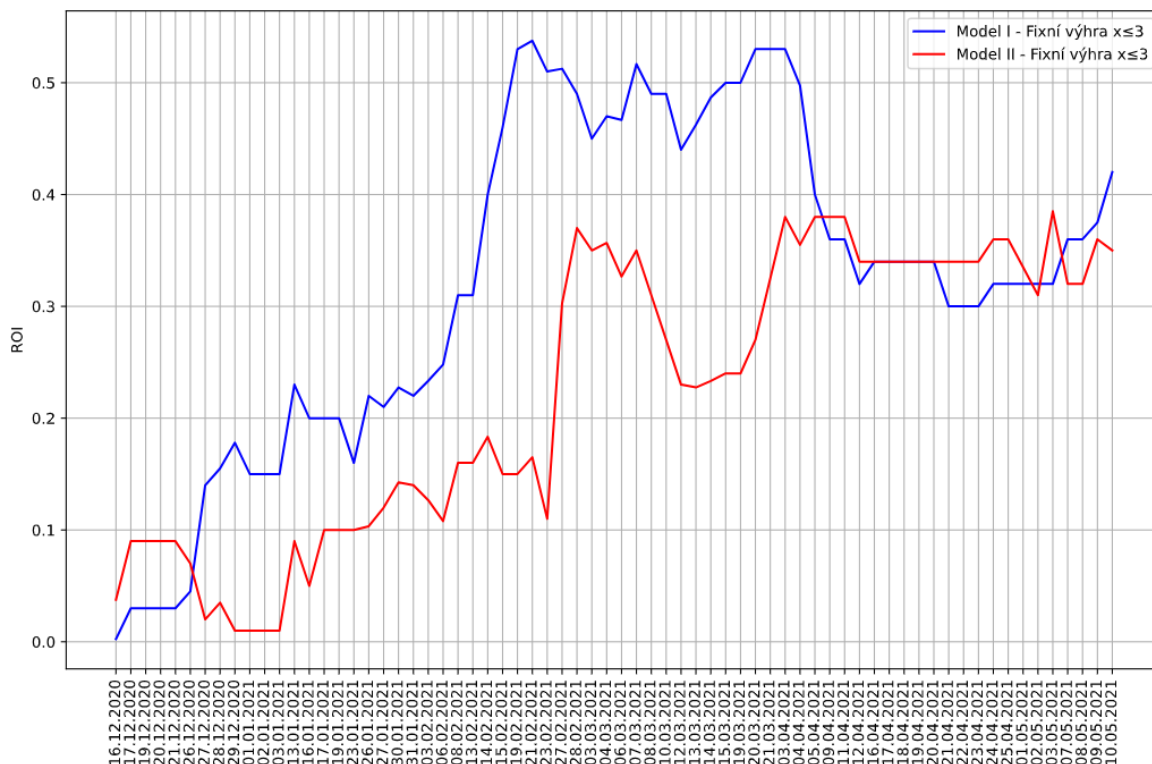
### 8.1 Porovnání sázkových strategií

*Tabulka 11 – Porovnání strategie fixní sázky*

Omezení	ROI <sub>Model I</sub>	ROI <sub>Model II</sub>
$x \leq 3$	0,42	0,35
$x \leq 2$	0,26	0,29
$x \leq 1,8$	0,35	0,19

V tabulce (tabulka 11) se porovnávají výsledky modelů pro sázkovou strategii fixních sázek. Na první pohled je vidět v tabulce, že oba modely pro všechna omezení dosahovaly v sezoně 2020/2021 Premier League zisku. Model I dosahoval vyššího zisku pro omezení  $x \leq 3$  a  $x \leq 1,8$ . Model II dosáhl vyššího zisku pouze pro omezení  $x \leq 2$ . Oba modely dosahovaly nejvyšší ziskovosti pro omezení  $x \leq 3$ .

Na obrázku (Obrázek 29) je zobrazeno porovnání průběhů pro omezení  $x \leq 3$ , u kterého dosahovaly oba modely nejvyšší hodnoty ROI na konci sezóny 2020/2021 anglické Premier League.



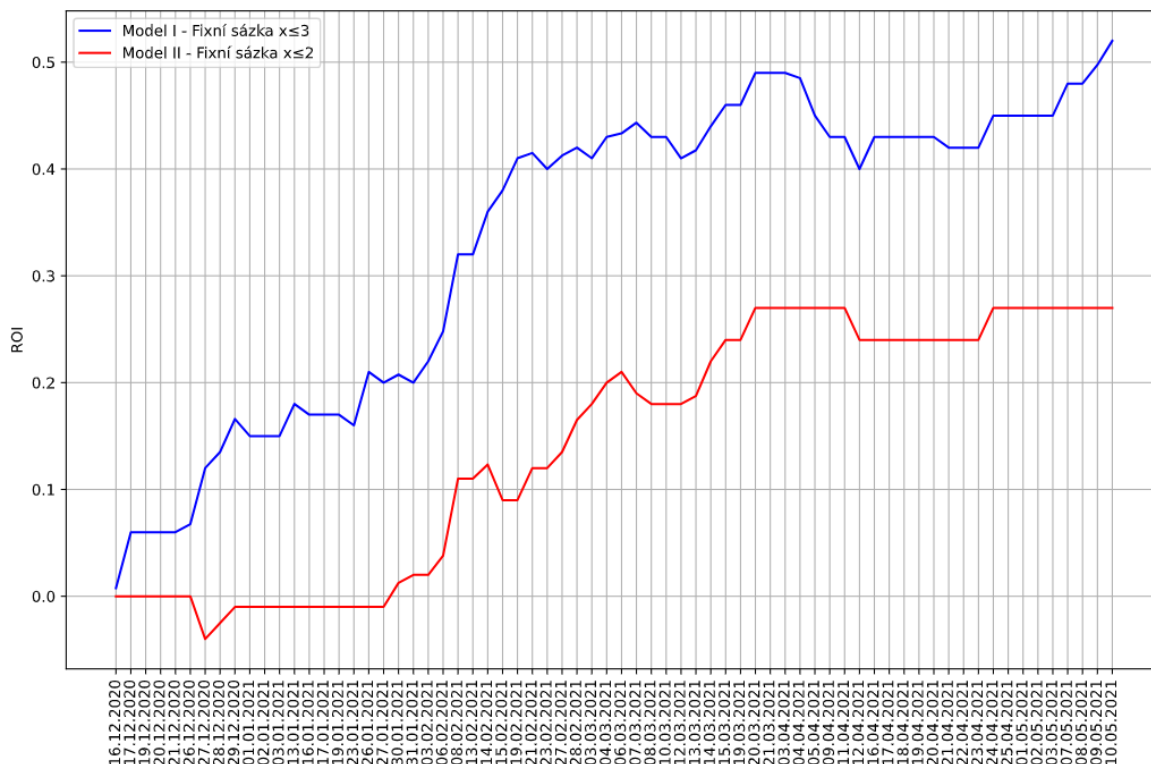
Obrázek 29 – Srovnání nejlepších výsledků modelů pro strategii fixních sázek

Tabulka 12 – Porovnání strategie fixní návratnosti

Omezení	Model I	Model II
$x \leq 3$	0,52	0,2
$x \leq 2$	0,39	0,27
$x \leq 1,8$	0,44	0,21

V tabulce (tabulka 12) se porovnávají výsledky modelů pro sázkovou strategii fixní návratnosti. Na první pohled je vidět, že oba modely pro všechna omezení dosahovaly v sezoně 2020/2021 Premier League zisku. Model I dosahoval vyššího zisku než model II pro všechna 3 omezení.

Na obrázku (Obrázek 30) je zobrazeno porovnání průběhů pro omezení  $x \leq 3$  u modelu I a omezení  $x \leq 2$  u modelu II. Pro tato omezení dosahovaly modely nejvyšší hodnoty ROI na konci sezóny 2020/2021 anglické Premier League.



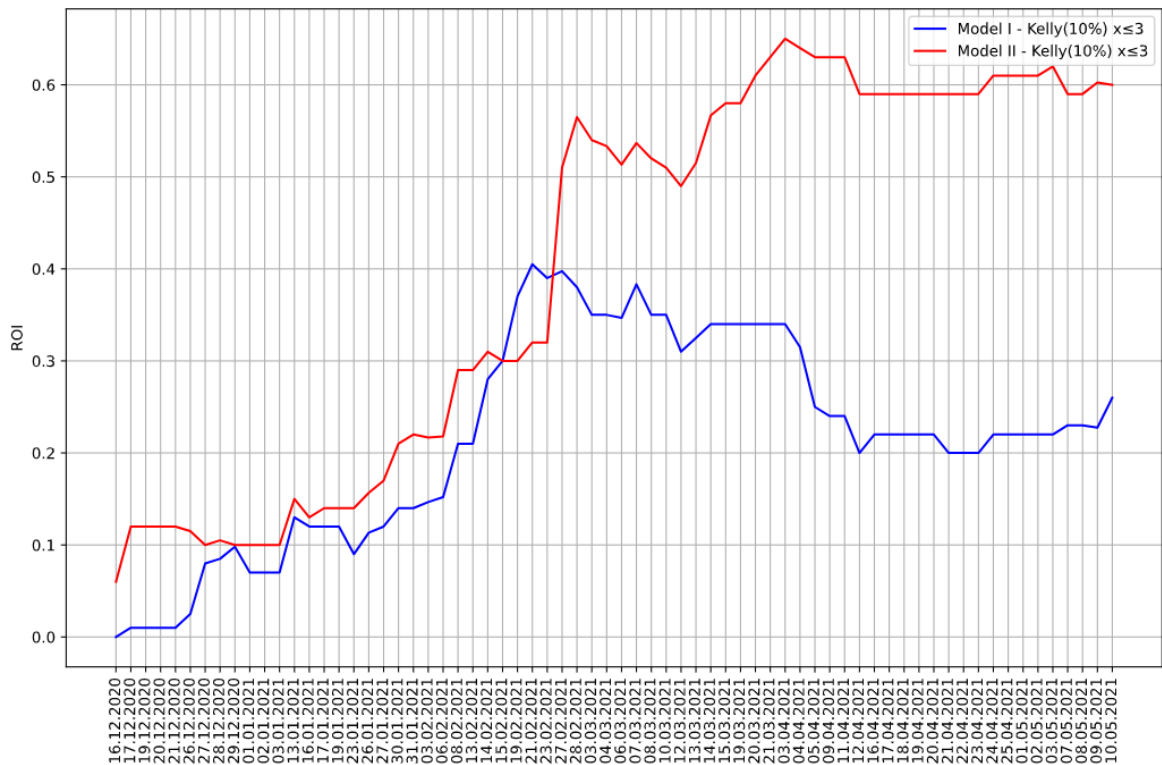
Obrázek 30 – Srovnání nejlepších výsledků modelů pro strategii fixní návratnosti

Tabulka 13 – Porovnání strategie 10 % Kellyho kritéria

Omezení	Model I	Model II
$x \leq 3$	0,26	0,6
$x \leq 2$	0,2	0,31
$x \leq 1,8$	0,22	0,25

V tabulce (tabulka 13) se porovnávají výsledky modelů pro sázkovou strategii 10 % Kellyho kritéria. Na první pohled je vidět, že oba modely pro všechna omezení dosahovaly na konci sezony 2020/2021 Premier League zisku. Model II dosahoval vyššího zisku než model I pro všechna 3 omezení.

Na obrázku (Obrázek 31) je zobrazeno porovnání průběhů pro omezení  $x \leq 3$ , u kterého dosahovaly oba modely nejvyšší hodnoty ROI na konci sezóny 2020/2021 anglické Premier League.



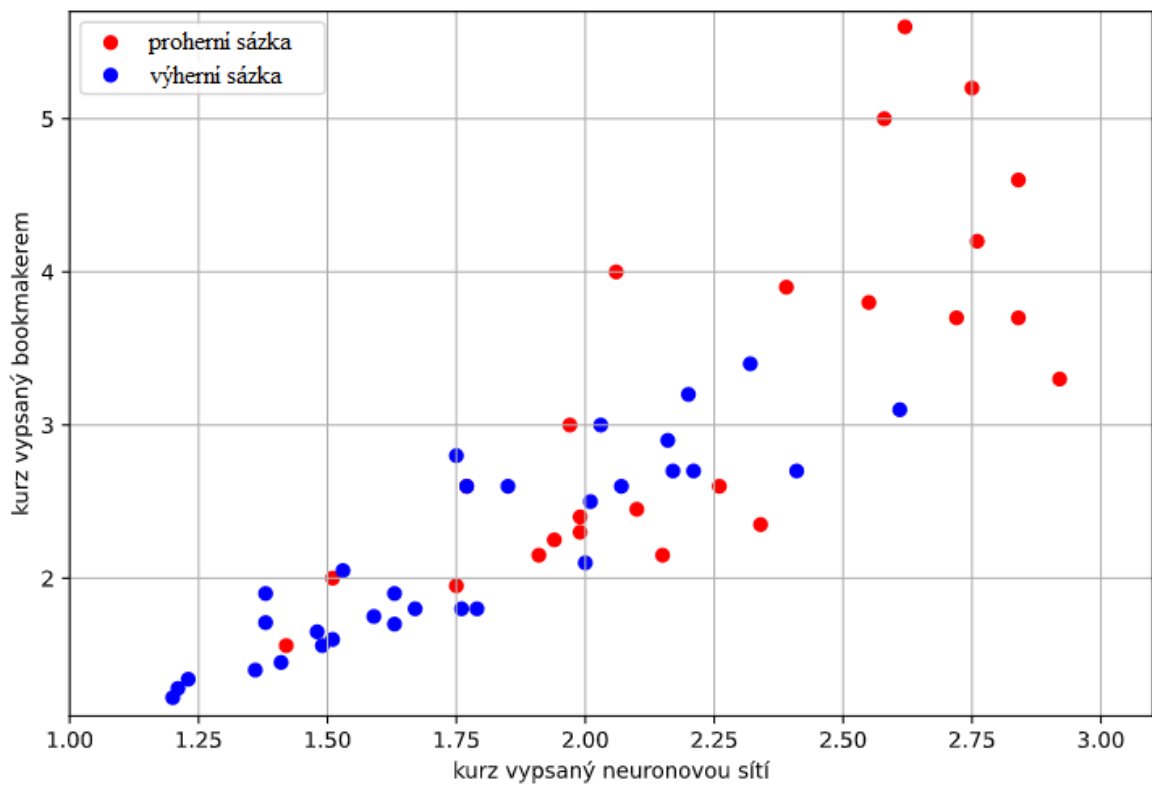
Obrázek 31 - Srovnání nejlepších výsledků modelů pro strategii 10 % Kellyho kritéria

## 8.2 Porovnání kurzů

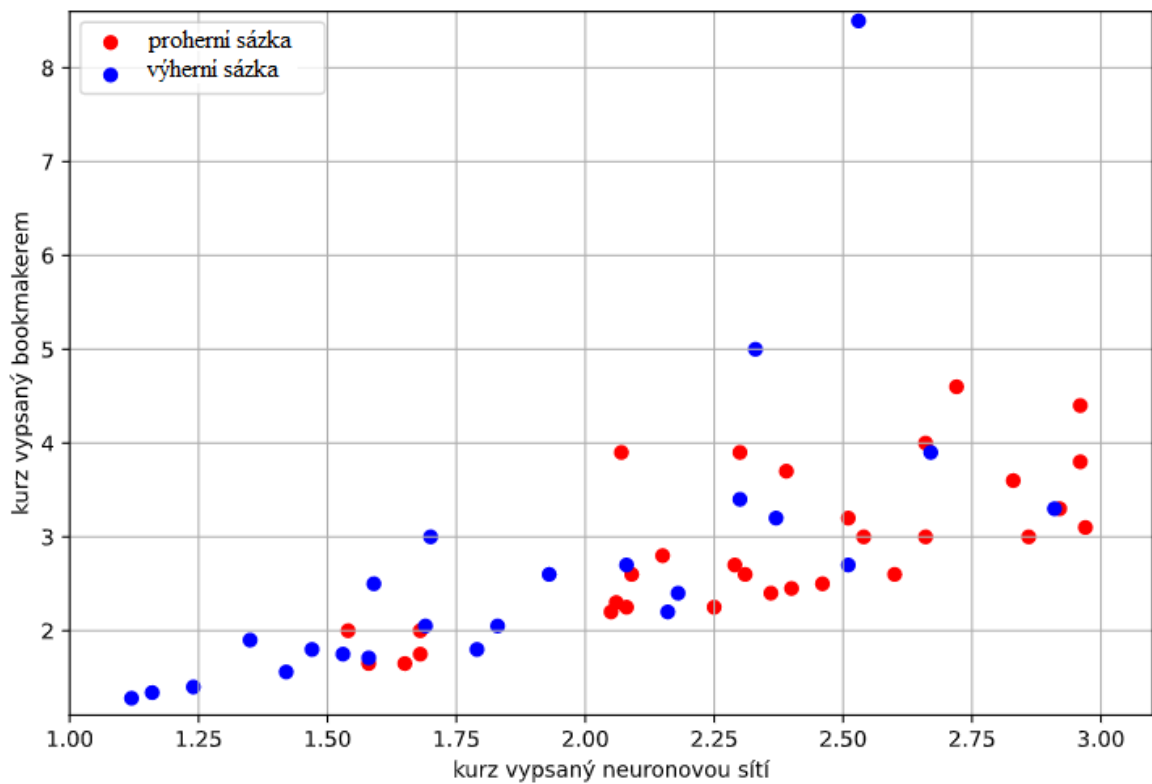
V této kapitole je pro hodnotné kurzy obou vytvořených modelů porovnán kurz sázkové kanceláře Tipsport s kurzem neuronové sítě. Grafy jsou vykresleny pro omezení  $x \leq 3$ , ale čtenář si může imaginárním posunutím x-ové osy, představit výsledky pro jakékoliv nižší omezení.

Na obrázku (Obrázek 32) je zobrazeno srovnání kurzů modelu I pro omezení  $x \leq 3$  a kurzu společnosti Tipsport. Od kurzu 2,5 vypsánoho neuronovou sítí by byla pouze jedna hodnotná sázka (kurz neuronové sítě je nižší než kurz bookmakera) výhodní, proto by v reálné aplikaci byly vhodné snížit přípustnou hranici u vypsánoho kurzu neuronovou sítí např. na  $x \leq 2,5$ .

Na obrázku (Obrázek 33) je obdobně zobrazeno srovnání kurzů pro modelu. Na první pohled lze zaznamenat významný rozdíl, kdy kurz u Tipsportu dosahoval hodnoty přes 8 a u modelu II byl kolem 2,5. I přesto že tahle sázka by měla pravděpodobnost úspěchu menší než 50 % (kurz 2,5 odpovídá pravděpodobnosti 40 %) byla by výhodní. Model II dosáhl lepších výsledků pro vyšší kurzy než model I, a proto byl tak velký rozdíl v jejich ziscích pro strategii 10 % Kellyho kritéria.



Obrázek 32 - Srovnání kurzů hodnotných sázek pro model I



Obrázek 33 - Srovnání kurzů hodnotných sázek pro model II

### 8.3 Porovnání přesností

Pro oba vytvořené modely byla vypočtena přesnost a srovnána v tabulce (Tabulka 10) s modelem sázkové společnosti Tipsport. Pro model společnosti Tipsport musely být přepočteny pravděpodobnosti pomocí webové aplikace [25], kvůli započítané marži, která dosahuje hodnoty 10 %. Pro každý model byla vykreslena matice záměn.

*Tabulka 14 – Srovnání přesností modelů*

	Model Tipsport [%]	Model I [%]	Model II
Přesnost	68,79	71,63	70,92

V tabulce (Tabulka 14) je vidět, že oba vytvořené modely dosáhly vyšší přesnosti než model sázkové společnosti Tipsport. Jelikož nejvyšší přesnosti 71,63 % dosáhl Model I, lze konstatovat, že lze vytvořit více přesné modely z dat, které jsou veřejně dostupné na internetu, než které používají sázkové společnosti.

*Tabulka 15 – Matice záměn pro model I*

		Výstup klasifikátoru		
		Výhra domácího týmu	Remíza	Výhra hostujícího týmu
Skutečnost	Výhra domácího týmu	64	0	17
	remíza	7	0	6
	Výhra hostujícího týmu	10	0	37

V tabulce (Tabulka 15) je vidět, že model I správně předpověděl výhru domácího týmu v 64 zápasech a výhru hostujícího mužstva ve 37 zápasech. Remízu se modelu I správně předpovědět nepodařilo.



Tabulka 16 – Matice záměn pro model II

		Výstup klasifikátoru		
		Výhra domácího týmu	Remíza	Výhra hostujícího týmu
Skutečnost	Výhra domácího týmu	57	0	17
	remíza	7	0	6
	Výhra hostujícího týmu	11	0	43

V tabulce (Tabulka 16) je vidět, že model I správně předpověděl výhru domácího týmu v 57 zápasech a výhru hostujícího mužstva ve 43 zápasech. Stejně jako modelu I se modelu II nepodařilo predikovat remízu ani v jednom případě.

Tabulka 17 – Matice záměn pro model sázkové kanceláře Tipsport

		Výstup klasifikátoru		
		Výhra domácího týmu	Remíza	Výhra hostujícího týmu
Skutečnost	Výhra domácího týmu	58	0	16
	remíza	5	0	8
	Výhra hostujícího týmu	15	0	39

V tabulce (Tabulka 17) je vidět, že model společnosti Tipsport správně předpověděl výhru domácího týmu v 58 případech, výhru hostujícího týmu ve 39 případech a ani jednou se mu nepodařilo správně predikovat remízu.

## ZÁVĚR

V teoretické části práce je uveden stručný úvod k problematice kurzového sázení a nutný teoretický výklad pro strojové učení, neuronové sítě a genetické algoritmy.

V praktické části práce je popsáno, jak byl vytvořen dataset, který je zkonstruován speciálně pro kurzové sázky o poločasové přestávce fotbalového utkání. Autorovi práce není známa veřejná existence podobně unikátního datasetu i přes podrobnou rešerši v dané oblasti. Dále byly vytvořeny 2 modely neuronových sítí v programovacím jazyku Python, pomocí modulu Keras. Pro oba modely je pomocí genetického algoritmu vybrána podmnožina atributů, která dosahovala maximální hodnoty fitness. Poté byly pro modely vytvořeny simulace 3 vybraných sázkových strategií. Pro každou sázkovou strategii byly definovány hranice a vykresleny průběhy návratnosti investic během sezony 2020/2021 Premier League. Následně pro každý model bylo vybráno nastavení neuronové sítě, které dosahovalo z 10 simulací nejvyšší přesnosti na validační množině. Pro tyto 2 vybrané modely byly pro hodnotné sázky mezi sebou porovnány vybrané sázkové strategie z hlediska návratnosti investice a také byl pro každý model vykreslen rozdíl kurzů mezi sázkovou kanceláří Tipsport a modelem neuronové sítě. Vytvořené modely byly také porovnány pomocí přesnosti se sázkovou kanceláří Tipsport. Oba modely v přesnosti poráží model sázkové kanceláře Tipsport. Nejlepší vytvořený model (model I) je o 2,84 % přesnější než model Tipsportu. Oba modely pro všechny 3 sázkové strategie ve všech provedených simulacích vykazovaly na konci sezony 2020/2021 anglické nejvyšší soutěže Premier League zisk.

Z výsledků diplomové práce autor práce konstatuje, že neuronové sítě jsou vhodným nástrojem pro porážení bookmakera v jeho vlastní hře a že je možné vytvořit modely, které budou dosahovat vyšší přesnosti než modely, které používají sázkové kanceláře. Důležité je také zmínit, že je možné vytvořit kvalitní datasety z veřejně dostupných dat na internetu, které pak mají lví podíl na úspěchu neuronových sítí.

V příloze jsou pro každý model detailně popsány použité atributy. Použité datasety a vytvořené modely ve frameworku Keras jsou dostupné na přiloženém CD.

Budoucí práce se bude týkat reálného použití modelu I na živých sázkových trzích v sezoně Premier League 2021/2022. Model I bude přetrénován tak, že validační sada rozšíří sadu trénovací a jako validační množina bude použita aktuální testovací sada, tedy letošní sezona Premier League 2020/2021. Dále chce autor práce vytvořit simulace i pro ostatní evropské soutěže, aby zjistil, jestli je možné pomocí modelu, který je natrénován pouze na anglické

lize generovat zisk i v jiných evropských soutěžích, případně zda je zapotřebí pro každou soutěž vytvořit speciální model i dataset.

**SEZNAM POUŽITÉ LITERATURY**

- [1] BUCHDAHL, Joseph. *Fixed Odds Sports Betting: Statistical Forecasting and Risk Management*. High Stakes, 2003. ISBN 978-1843440192.
- [2] *Tipsport* [online]. [cit. 2020-08-13]. Dostupné z: <https://www.tipsport.cz/napoveda>
- [3] Franck, Egon & Verbeek, Erwin & Nüesch, Stephan. (2012). *Inter-Market Arbitrage in Betting. Economica*. 80. 10.2139/ssrn.1503375.
- [4] Nyquist, R., & Pettersson, D. (2017). Football match prediction using deep learning.
- [5] BORØY-JOHNSEN, Simon. *Beating the bookmakers - Using artificial neural networks to profit from football betting*. 2017. NTNU.
- [6] Samba, Stefan. (2019). *Football Result Prediction by Deep Learning Algorithms*. 10.13140/RG.2.2.25014.45122.
- [7] Tax, Niek & Joustra, Yme. (2015). *Predicting The Dutch Football Competition Using Public Data: A Machine Learning Approach*. 10.13140/RG.2.1.1383.4729.
- [8] Godin, F., Zuallaert, J., Vandersmissen, B., De Neve, W., and Van de Walle, R. (2014). *Beating the bookmakers:*
- [9] BURKOV, Andriy. *The Hundred-Page Machine Learning Book*. 2019. ISBN 978-1999579517.
- [10] CHOLLET, François. *Deep learning with Python*. Shelter Island, NY: Manning, [2018]. ISBN 978-161-7294-433.
- [11] GÉRON, A. *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. . Sebastopol, CA: O'Reilly Media, 2017. ISBN 978-1491962299.
- [12] GRUS, Joel. *Data science from scratch: first principles with Python*. Second edition. Beijing: O'Reilly, 2019. ISBN 978-149-2041-139.
- [13] *File:Artificial neural network.png* [online]. In: . [cit. 2020-08-14]. Dostupné z: [https://commons.wikimedia.org/wiki/File:Artificial\\_neural\\_network.png](https://commons.wikimedia.org/wiki/File:Artificial_neural_network.png)
- [14] GOODFELLOW, Ian, Yoshua BENGIO a Aaron COURVILLE. *Deep learning*. Cambridge: MIT Press, [2016]. Adaptive computation and machine learning (MIT Press). ISBN 978-0262035613.

- [15] VOLNÁ, Eva. *Neuronové sítě a genetické algoritmy*. Ostrava: Ostravská univerzita, 1998. ISBN 80-704-2762-0.
- [16] *File:MultiLayerPerceptron.png* [online]. In: . [cit. 2020-08-14]. Dostupné z: <https://commons.wikimedia.org/wiki/File:MultiLayerPerceptron.png>
- [17] GOLDBERG, David E. *Genetic algorithms in search. Optimization, and Machine Learning*, 1989.
- [18] *File:Computational.science.Genetic.algorithm.Crossover.Two.Point.svg* [online]. In: . [cit. 2020-05-20]. Dostupné z: <https://upload.wikimedia.org/wikipedia/commons/4/47/Computational.science.Genetic.algorithm.Crossover.Two.Point.svg>
- [19] *Whoscored* [online]. [cit. 2020-08-14]. Dostupné z: <https://www.whoscored.com/>
- [20] *Livesport* [online]. [cit. 2020-05-20]. Dostupné z: <https://www.livesport.cz/>
- [21] *Beautiful Soup* [online]. [cit. 2020-08-14]. Dostupné z: <https://www.crummy.com/software/BeautifulSoup/>
- [22] *Keras* [online]. [cit. 2020-05-20]. Dostupné z: <https://www.keras.io/>
- [23] *TensorFlow* [online]. [cit. 2020-05-20]. Dostupné z: <https://www.tensorflow.org/>
- [24] *FeatureSelectionGA* [online]. [cit. 2020-05-20]. Dostupné z: <https://pypi.org/project/feature-selection-ga/>
- [25] KELLY JR, John L. A new interpretation of information rate. In: *The Kelly Capital Growth Investment Criterion: Theory and Practice*. 2011. p. 25-34.
- [26] *Margin Calculator* [online]. [cit. 2020-05-20]. Dostupné z: <https://www.pinnacle.com/en/betting-resources/betting-tools/margin-calculator>

**SEZNAM POUŽITÝCH ZKRATEK**

MLP Vícevrstvý perceptron (z angl. Multi-layer perceptron)

ReLU Rektifikovaná lineární jednotka (z angl. Rectified linear unit)

TF Přenosová funkce (z angl. Transfer function)

tanh Hyperbolický tangens

## SEZNAM OBRÁZKŮ

<i>Obrázek 1 – Strojové učení jako programové paradigma [10]</i> .....	15
<i>Obrázek 2 – Matematický model neuronu [13]</i> .....	18
<i>Obrázek 3 – Aktivační funkce</i> .....	19
<i>Obrázek 4 – Schéma vícevrstvého perceptronu [16]</i> .....	20
<i>Obrázek 5 – Dvoubodové křížení [18]</i> .....	23
<i>Obrázek 6 – Ukázka mutace jedince</i> .....	23
<i>Obrázek 7 – Data pro jednotlivá utkání na webu WhoScored.com</i> .....	26
<i>Obrázek 8 – Kurzy na Livesport.cz</i> .....	26
<i>Obrázek 9 – Live kurzy na Tipsport.cz</i> .....	27
<i>Obrázek 10 – Namapování chromozomu na vektor atributů</i> .....	27
<i>Obrázek 11 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky (<math>x \leq 3</math>)</i> .....	34
<i>Obrázek 12 - Porovnání průběhů ROI pro sazkovou strategii fixní sázky (<math>x \leq 2</math>)</i> .....	34
<i>Obrázek 13 - Porovnání průběhů ROI pro sazkovou strategii fixní sázky (<math>x \leq 1,8</math>)</i> .....	35
<i>Obrázek 14 - Porovnání průběhů ROI pro sazkovou strategii fixní výhry (<math>x \leq 3</math>)</i> .....	36
<i>Obrázek 15 - Porovnání průběhů ROI pro sazkovou strategii fixní sázky (<math>x \leq 2</math>)</i> .....	37
<i>Obrázek 16 - Porovnání průběhů ROI pro sazkovou strategii fixní sázky (<math>x \leq 1,8</math>)</i> .....	37
<i>Obrázek 17- Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria</i> .....	39
<i>Obrázek 18 - Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria</i> .....	39
<i>Obrázek 19 - Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria</i> .....	40
<i>Obrázek 20 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky (<math>x \leq 3</math>)</i> .....	42
<i>Obrázek 21 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky (<math>x \leq 2</math>)</i> .....	43
<i>Obrázek 22 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky (<math>x \leq 1,8</math>)</i> .....	43
<i>Obrázek 23 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky (<math>x \leq 3</math>)</i> .....	45
<i>Obrázek 24 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky (<math>x \leq 2</math>)</i> .....	45
<i>Obrázek 25 – Porovnání průběhů ROI pro sazkovou strategii fixní sázky (<math>x \leq 1,8</math>)</i> .....	46
<i>Obrázek 26 - Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria</i> .....	48
<i>Obrázek 27 - Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria</i> .....	49
<i>Obrázek 28 - Porovnání průběhů ROI pro sazkovou strategii 10 % Kellyho kritéria</i> .....	50
<i>Obrázek 29 – Srovnání nejlepších výsledků modelů pro strategii fixních sázek</i> .....	52
<i>Obrázek 30 – Srovnání nejlepších výsledků modelů pro strategii fixní návratnosti</i> .....	53
<i>Obrázek 31 - Srovnání nejlepších výsledků modelů pro strategii 10 % Kellyho kritéria</i> ....	54
<i>Obrázek 32 - Srovnání kurzů hodnotných sázek pro model I</i> .....	55
<i>Obrázek 33 - Srovnání kurzů hodnotných sázek pro model II</i> .....	55

**SEZNAM TABULEK**

<i>Tabulka 1 – Ukázka arbitrážního sázení .....</i>	13
<i>Tabulka 2 – Porovnání Kellyho kritéria .....</i>	29
<i>Tabulka 3 – Použité parametry Modelu I .....</i>	32
<i>Tabulka 4 – Porovnání pro sázkovou strategii fixní sázka .....</i>	33
<i>Tabulka 5 - Porovnání pro sázkovou strategii fixní výhry.....</i>	35
<i>Tabulka 6 - Porovnání pro sázkovou strategii 10 % Kellyho kritéria .....</i>	38
<i>Tabulka 7 – Použité parametry Modelu II.....</i>	40
<i>Tabulka 8 - Porovnání pro sázkovou strategii fixní sázka.....</i>	41
<i>Tabulka 9 - Porovnání pro sázkovou strategii fixní výhra .....</i>	44
<i>Tabulka 10 - Porovnání pro sázkovou strategii 10 % Kellyho kritéria.....</i>	46
<i>Tabulka 11 – Porovnání strategie fixní sázky.....</i>	51
<i>Tabulka 12 – Porovnání strategie fixní návratnosti .....</i>	52
<i>Tabulka 13 – Porovnání strategie 10 % Kellyho kritéria.....</i>	53
<i>Tabulka 14 – Srovnání přesnosti modelů .....</i>	56
<i>Tabulka 15 – Matice záměn pro model I .....</i>	56
<i>Tabulka 16 – Matice záměn pro model II.....</i>	57
<i>Tabulka 17 – Matice záměn pro model sázkové kanceláře Tipsport .....</i>	57
<i>Tabulka 18 – Atributy modelu I .....</i>	66
<i>Tabulka 19 - Atributy modelu II .....</i>	69



## **SEZNAM PŘÍLOH**

Příloha P I: Popis atributů modelu I

Příloha P II: Popis atributů modelu II

Příloha P III: Elektronické přílohy na CD

## PŘÍLOHA P I: POPIS ATRIBUTŮ MODELU I

Tabulka 18 – Atributy modelu I

ratings_home_1_to_HT	hodnocení domácího týmu v prvním poločase utkání
ratings_away_1_to_HT	hodnocení hostujícího týmu v prvním poločase utkání
yellow_cards_home_1_to_HT	počet žlutých karet domácího týmu v prvním poločase utkání
yellow_cards_away_1_to_HT	počet žlutých karet hostujícího týmu v prvním poločase utkání
goals_home_1_to_HT	počet vstřelených branek domácího týmu v prvním poločase utkání
goals_away_1_to_HT	počet vstřelených branek hostujícího týmu v prvním poločase utkání
woodwork_home_1_to_HT	počet nastřelených břevn domácího týmu v prvním poločase utkání
woodwork_away_1_to_HT	počet nastřelených břevn domácího týmu v prvním poločase utkání
shots_on_target_home_1_to_HT	počet střel na bránu domácího týmu v prvním poločase utkání
shots_on_target_away_1_to_HT	počet střel na bránu hostujícího týmu v prvním poločase utkání
shots_off_target_home_1_to_HT	počet střel mimo prostor branky domácího týmu v prvním poločase utkání
shots_off_target_away_1_to_HT	počet střel mimo prostor branky hostujícího týmu v prvním poločase utkání
shots_blocked_home_1_to_HT	počet zablokovaných střel domácího týmu v prvním poločase utkání
shots_blocked_away_1_to_HT	počet zablokovaných střel hostujícího týmu v prvním poločase utkání
touches_home_1_to_HT	počet doteků s míčem domácího týmu v prvním poločase utkání
touches_away_1_to_HT	počet doteků s míčem hostujícího týmu v prvním poločase utkání
pass_success_home_1_to_HT	počet úspěšných přihrávek domácího týmu v prvním poločase utkání
pass_success_away_1_to_HT	počet úspěšných přihrávek hostujícího týmu v prvním poločase utkání
total_passes_home_1_to_HT	celkový počet přihrávek domácího týmu v prvním poločase utkání
total_passes_away_1_to_HT	celkový počet přihrávek hostujícího týmu v prvním poločase utkání
accurate_passes_home_1_to_HT	počet úspěšných přihrávek domácího týmu v prvním poločase utkání
accurate_passes_away_1_to_HT	počet úspěšných přihrávek hostujícího týmu v prvním poločase utkání
key_passes_home_1_to_HT	Počet klíčových přihrávek domácího týmu v prvním poločase utkání

key_passes_away_1_to_HT	Počet klíčových přihrávek hostujícího týmu v prvním poločase utkání
dribble_success_home_1_to_HT	Počet úspěšných driblingů domácího týmu v prvním poločase utkání
dribble_success_away_1_to_HT	Počet úspěšných driblingů hostujícího týmu v prvním poločase utkání
interceptions_home_1_to_HT	Počet zachycených přihrávek hostujícího týmu v prvním poločase utkání
interceptions_away_1_to_HT	Počet zachycených přihrávek domácího týmu v prvním poločase utkání
corners_home_1_to_HT	Počet zahrávaných rohů domácího týmu v prvním poločase utkání
corners_away_1_to_HT	Počet zahrávaných rohů hostujícího týmu v prvním poločase utkání
dispossessed_home_1_to_HT	Počet odebraných míčů domácím týmem v prvním poločase utkání
dispossessed_away_1_to_HT	Počet odebraných míčů hostujícím týmem v prvním poločase utkání
h_saves_HT	Počet zákroků brankáře domácího týmu v prvním poločase utkání
a_saves_HT	Počet zákroků brankáře hostujícího týmu v prvním poločase utkání
h_cross_HT	Počet centrovaných míčů zahráných do před brankového prostoru zahráných domácím týmem v prvním poločase utkání
a_cross_HT	Počet centrovaných míčů zahráných do před brankového prostoru zahráných hostujícím týmem v prvním poločase utkání
h_short_HT	Počet krátkých přihrávek domácího týmu v prvním poločase utkání
a_short_HT	Počet krátkých přihrávek hostujícího týmu v prvním poločase utkání
h_passes_defensive_third	Počet přihrávek domácího týmu směřujících do obrané třetiny hřiště v prvním poločase utkání
a_passes_defensive_third	Počet přihrávek hostujícího týmu směřujících do obrané třetiny hřiště v prvním poločase utkání
h_passes_final_third	Počet přihrávek domácího týmu směřujících do útočné třetiny hřiště v prvním poločase utkání
a_passes_final_third	Počet přihrávek hostujícího týmu směřujících do útočné třetiny hřiště v prvním poločase utkání
h_odds	Průměrný kurz vypsany bookmakery na vítězství domácího týmu v utkání (před začátkem utkání)
d_odds	Průměrný kurz vypsany bookmakery na remízu v utkání (před začátkem utkání)

c	Predikovaná hodnota - Více rohů v zápasu : Tým, který zahraje více rohových kopů v utkání (H – domácí tým, D – remíza, A – hostující tým)
---	--

## PŘÍLOHA P II: POPIS ATRIBUTŮ MODELU II

Tabulka 19 - Atributy modelu II

d_odds	Průměrný kurz vypsáný bookmakery na remízu v utkání (před začátkem utkání)
yellow_cards_1_to_HT	Rozdíl v počtu obdržených žlutých karet mezi domácím a hostujícím týmem (v prvním poločase utkání)
goals_1_to_HT	Rozdíl v počtu vstřelených branek mezi domácím a hostujícím týmem (v prvním poločase utkání)
total_shots_1_to_H	Rozdíl v celkovém počtu střel mezi domácím a hostujícím týmem (v prvním poločase utkání)
woodwork_1_to_HT	Rozdíl v počtu nastřelených břevien mezi domácím a hostujícím týmem (v prvním poločase utkání)
shots_on_target1_to_HT	Rozdíl v počtu střel na bránu mezi domácím a hostujícím týmem (v prvním poločase utkání)
total_passes_1_to_HT	Rozdíl v počtu zahráných přihrávek mezi domácím a hostujícím týmem (v prvním poločase utkání)
dribbles_attempted_1_to_HT	Rozdíl v počtu pokusů o dribling mezi domácím a hostujícím týmem (v prvním poločase utkání)
dribble_success_1_to_HT	Rozdíl v počtu úspěšných driblingů mezi domácím a hostujícím týmem (v prvním poločase utkání)
offensive_aerials_1_to_HT	Rozdíl v počtu vyhraných útočných soubojů mezi domácím a hostujícím týmem (v prvním poločase utkání)
tackles_1_to_HT	Rozdíl v počtu zastavení soupeře mezi domácím a hostujícím týmem (v prvním poločase utkání)
clearances_1_to_HT	Rozdíl v počtu odehraní míče z nebezpečného prostoru před brankou mezi domácím a hostujícím týmem (v prvním poločase utkání)
interceptions_1_to_HT	Rozdíl v počtu zachycených přihrávek mezi domácím a hostujícím týmem (v prvním poločase utkání)
fouls_1_to_HT	Rozdíl v počtu odpískaných faulů mezi domácím a hostujícím týmem (v prvním poločase utkání)
saves_HT	Rozdíl v počtu zákroků mezi domácím a hostujícím brankářem (v prvním poločase utkání)

6yard_HT	Rozdíl v počtu střel, které byly zahrávány z malého vápna mezi domácím a hostujícím týmem (v prvním poločase utkání)
cross_HT	Rozdíl v počtu centrovaných míčů do před brankového prostoru mezi domácím a hostujícím týmem (v prvním poločase utkání)
free_kick_HT	Rozdíl v počtu zahrávání standartních situací mezi domácím a hostujícím týmem (v prvním poločase utkání)
throw_in_HT	Rozdíl v počtu autových vhazování mezi domácím a hostujícím týmem (v prvním poločase utkání)
h_chipped_HT	Rozdíl v počtu přihrávek, které letěly vzduchem mezi domácím a hostujícím týmem (v prvním poločase utkání)
h_ground_HT	Rozdíl počtu přihrávek zahranych po zemi mezi domácím a hostujícím týmem (v prvním poločase utkání)
h_dribbles	Rozdíl v počtu driblinků mezi domácím a hostujícím týmem (v prvním poločase utkání)
h_blocks	Rozdíl v počtu zblokovaných střel a zblokovaných centrovaných míčů mezi domácím a hostujícím týmem (v prvním poločase utkání)
c	Predikovaná hodnota - Více rohů v zápasu : Tým, který zahraje více rohových kopů v utkání (H – domácí tým, D – remíza, A – hostující tým)