

UNIVERZITA KARLOVA  
Fakulta tělesné výchovy a sportu

## **Testování modelu pro predikci výsledků tenisových utkání**

Diplomová práce

Vedoucí diplomové práce:

**Doc. PhDr. Jan Šíma, Ph.D.**

Vypracoval:

**Bc. Daniel Karas**

Praha, prosinec 2022

Prohlašuji, že jsem závěrečnou diplomovou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, dne

.....

podpis diplomanta

## **Evidenční list**

Souhlasím se zapůjčením své diplomové práce ke studijním účelům. Uživatel svým podpisem stvrzuje, že tuto diplomovou práci použil ke studiu a prohlašuje, že ji uvede mezi použitými prameny.

Jméno a příjmení:

Fakulta / katedra:

Datum vypůjčení:

Podpis:

---

## **Poděkování**

Rád bych poděkoval panu doc. PhDr. Janu Šímovi, Ph.D., vedoucímu mé diplomové práce, za odborné vedení, cenné rady a vstřícnost při konzultacích.

## **Abstrakt**

**Název:** Testování modelu pro predikci výsledků tenisových utkání

**Cíle:** Hlavním cílem této práce je vytvořit jednoduchý a snadno pochopitelný model predikce výsledků tenisového utkání.

**Metody:** Pro získání relevantních a aktuálních informací byla v této diplomové práci využita analýza sekundárních dat. V práci byl kladen důraz na to, aby data byla validní a pocházela z důvěryhodných zdrojů. Dále byly využity matematické a statistické metody k získání výpočtů pro určení správných predikcí. Získaná data byla zpracována a poté prezentována prostřednictvím tabulek.

**Výsledky:** Upravený model založený na Low-point modelu, ale zjednodušený tak, aby mu rozuměla široká veřejnost, a který využíval snadno dohledatelná a ihned dostupná data, úspěšně predikoval turnaj Australian Open 2021. Na konci turnaje jsme měli na fiktivním účtě více peněz než na začátku. Model dokázal správně predikovat 5 ze 7 zápasů a byl v zisku při sázce na predikovaného vítěze i na sázku na výhodnost kurzu. Při ověření modelu na dalších dvou grandslamech byl však tento model ve ztrátě, a nemůžeme ho tak označit za dlouhodobě výnosný.

**Klíčová slova:** tenis, predikce, sázka, Australian Open, grandslam, výhodnost,

## **Abstract**

**Title:** Testing of model for predicting the outcome of professional tennis matches

**Objectives:** The aim of this work is to create a simple and easy to understand model to predict tennis matches.

**Methods:** In this thesis it was used secondary analysis of the data for gaining a relevant and current information. In this work it was emphasized on valid data, which should come from the reliable sources. Further, mathematic and statistical methods were used to obtain calculation for determination correct predictions. The obtained data were processed a then they were presented within tables.

### **Results:**

The modified model based on the Low-point model, which was simplified so the general public could understand it and which used easily searchable and immediately available data, successfully predicted the tournament Australian Open 2021. At the end of the tournament we had more money on the fictitious account than at the beginning. The model was able to correctly predict 5 out of 7 matches and it was in profit when betting on the predicted winner as well as betting on the advantage of the odds. However, when the model was validated at the next two Grand Slams, this model was at loss, so we cannot designate it as long-term profitable.

**Keywords:** tennis, prediction, bet, Australian Open, Grand Slam

# Obsah

1	ÚVOD.....	12
2	TEORETICKÁ VÝCHODISKA PRÁCE.....	14
2.1	Tenis.....	14
2.1.1	Hrací povrch.....	14
2.1.2	Bodový systém.....	15
2.1.3	Bodování ve hře.....	15
2.1.4	Pořadí podání.....	15
2.1.5	Tenisové turnaje.....	16
2.2	Oficiální bodovací systém hráčů.....	17
2.3	Australian Open.....	18
2.4	Finanční odměny v turnaji.....	18
2.5	Zákon o hazardních hrách.....	19
2.6	Predikce.....	20
2.7	Sázka.....	21
2.8	Výhodnost sázky.....	22
2.9	Převod kurzů na pravděpodobnosti.....	22
2.10	Strategie sázení.....	23
2.10.1	Strategie stejný vklad.....	24
2.10.2	Strategie vsadit vše – all in.....	24
2.10.3	Strategie Martingale.....	24
2.10.4	Strategie Fibonacci.....	25
2.11	Data získaná ze statistik.....	25
2.12	Markovovy řetězce.....	30
2.13	Metody predikce tenisových utkání.....	31
2.13.1	Win / Loss index.....	31
2.13.2	Head to head analýza.....	32

2.13.3	Magnus-Klaassen model.....	32
2.13.4	Barnett-Clarke model.....	33
2.13.5	Baseline Point Model.....	34
2.13.6	Corral-Prieto-Rodriguez model .....	35
2.13.7	Point model.....	36
2.13.8	Vanilla Bradley-Terry model.....	37
2.13.9	Free Parameter Point model.....	37
2.13.10	Low-level point model .....	38
3	CÍLE A ÚKOLY PRÁCE.....	42
3.1	Cíl výzkumného projektu.....	42
4	METODIKA PRÁCE .....	44
4.1	Základní soubor.....	44
4.2	Výběrový soubor.....	44
4.3	Využití modelu Low-level point .....	45
4.4	Sběr dat.....	45
4.5	Pavouk turnaje.....	46
5	VÝSLEDKY .....	49
5.1	Djokovič v zápase s Chardym.....	49
5.2	Čtvrtfinálové utkání Nadal vs. Tsitsipas .....	50
5.3	Statistiky zápasu.....	51
5.4	Celkový přehled všech analyzovaných výsledků Australian Open 2021 .....	59
5.5	Výsledek predikce dalších vybraných grandslamů .....	61
5.5.1	Australian Open 2022 .....	63
5.5.2	United States Open 2022 .....	65
6	DISKUZE .....	67
6.1	Porovnání výsledků modelu a reality na Australian Open 2021 .....	67
6.2	Porovnání pravděpodobností modelu a sázkových kanceláří .....	70



7 ZÁVĚR.....	72
SEZNAM LITERATURY.....	75
SEZNAM PŘÍLOH.....	80

## **Seznam použitých symbolů a zkratk**

AO2021: Australian Open 2021

AO2022: Australian Open 2022

ATP: Asociace tenisových profesionálů (z anglického „Association of Tennis Professionals“)

ITF: Mezinárodní tenisová federace (z anglického „International Tennis Federation“)

US2022: United States Open Tennis Championships 2022

US Open: United States Open Tennis Championships

# ÚVOD

Ve své práci se zabývám predikcí výsledku tenisového utkání. Mým cílem je prozkoumat existující modely pro predikci a jednu vybranou metodu zjednodušit tak, aby byla srozumitelná široké veřejnosti, jednoduše se do výpočtů vložila data, která budou lehce dohledatelná a vyzkoušet ji na konkrétním turnaji.

V dnešní době jsou veřejně přístupné webové stránky, které sledují různé statistiky hráčů. Na základě těchto statistik vznikají stále propracovanější modely predikce. Tato práce obsahuje rešerši nejvíce využívaných modelů, které se zabývají určováním výsledků tenisových utkání. Pro tuto práci je použita metoda založená na Low-Point modelu. Metoda bude používat stejné výpočty a vzorce, s tím rozdílem, že do něj budou vstupovat pouze data z daného jednoho turnaje od prvního kola.

Vybraná metoda je aplikovaná na tenisový turnaj Australian Open 2021. Tento turnaj byl zvolen, protože velké množství turnajů bylo na základě hygienických opatření ohledně COVID-19 zrušeno nebo posunuto na pozdější dobu. Australian Open se odehrál ve dnech 8. až 21. února 2021 v Melbourne parku. Tento turnaj je součástí Grandslamu, v němž je vysoká odměna za postup do dalšího kola, a tak se předpokládá, že hráči budou více motivováni zápas vyhrát a pravděpodobnost sázkového podvodu bude tedy minimální.

Pro zvolený model slouží jako vstupní data herní statistiky hráčů. K co nejpřesnější predikci výsledku turnaje je také nutné vyřešit, jak mít co nejaktuálnější a nejpřesnější data. Ty musí reflektovat aktuální formu tenisty, prostředí, ve kterém se hraje, povrch hrací plochy aj. Do utkání také vstupuje spousta psychologických vlivů jednotlivých hráčů, jako je stabilita, soutěživost, ambicióznost, agresivita, adaptabilita, sebevědomí, svědomitost atp. Tyto faktory však můžeme považovat spíše za doplňující, jelikož můžeme předpokládat, že na úrovni špičkového sportu se sportovci psychologickými vlivy nenechají příliš ovlivnit.

Ověření vybrané metody bylo provedeno proti statistikám odehraných zápasů. Po odehrání zápasu se hodnoty vložily do tabulky s hodnotami predikce a vyjádřila se procentuální odchylka modelu od reality, abychom zjistili, jak se reálná data liší od předpovídaných.

Výsledky byly také porovnány s kurzy sázkové kanceláře a bylo zkoumáno, zda nám model dokáže ve vybraných turnajích při stejné výši sázky generovat zisk. To znamená,

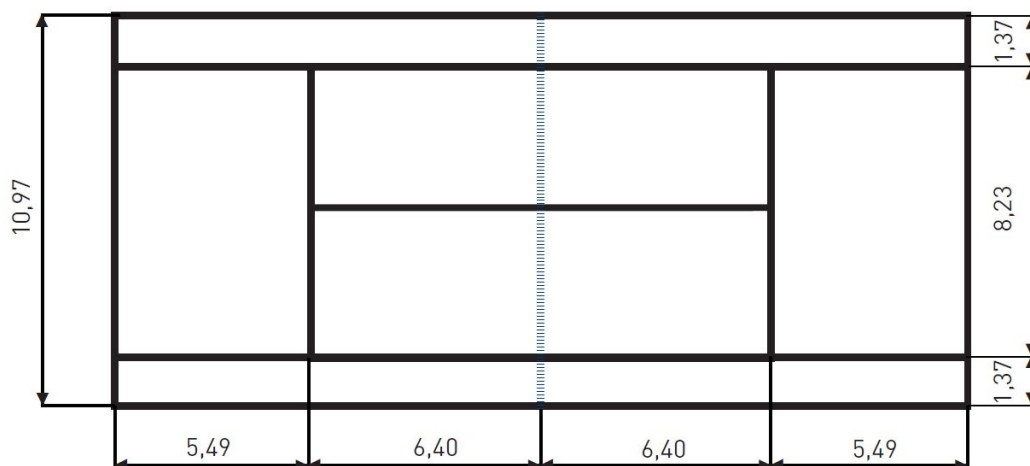
že jsme na každý zápas hypoteticky vsadili stejnou částku a podle kurzů a úspěšnosti predikce spočítali, zda se nám tento model vyplatí. Mým cílem není podpora gamblerství, ale poskytnutí dalších dat pro rozšíření obzoru sázejících. Na sportovní utkání a výsledky se sázelo již od nepaměti a jedná se bezesporu o velmi rozšířený byznys, který v naší společnosti bude přetrvávat i na dále. Tento model by mohl být pomocí pro sázkaře, aby byli schopni objektivně posoudit všechny proměnné a na základě toho se rozhodnout, na kterého hráče vsadit své peníze. Cílem tohoto modelu je tedy co nejjednodušší sběr dat a vytvoření systému výpočtů, díky kterému by si každý, kdo má základní znalosti matematiky a práce v tabulkách, mohl tyto výpočty udělat sám.

# TEORETICKÁ VÝCHODISKA PRÁCE

V této kapitole se budeme zabývat tenisem jako hrou a jeho pravidly, jejichž znalost je pro budoucí výpočty zcela nezbytná. K pochopení výpočtů a vzorců je například důležité znát systém střídání podání ve hře a počítání vítězných míčků, což nám pomůže pochopit Markovovy řetězce popsané v kapitole 2.8. a následně další modely použité v této kapitole.

## 1.1 Tenis

Tenis je raketová hra, která se hraje na tenisovém hřišti obdélníkového tvaru. Hřiště je dlouhé 23,77 m a 8,23 m široké. Tento obdélník je v polovině rozdělen tenisovou sítí ve výšce 1,07 m. Každý hráč hraje na jedné polovině hřiště. To znamená, že tenis je bezkontaktní hra. Cílem hráče je odpálit raketou míč na soupeřovu polovinu tak, aby soupeř nebyl schopný míček odehrát zpět a získat tak bod (ATP Tour, Inc., 2021; ITF LTD, 2021) (ATP Tour, Inc., 2021).



Obrázek 1 Velikost tenisového kurtu v metrech

### 1.1.1 Hrací povrch

Krásou tenisové hry spočívá i v různorodosti povrchů kurtu. Každému hráči vyhovuje jiný povrch, a proto mohou být zápasy zajímavé, i pokud proti sobě nastoupí hráči, kteří spolu již hráli, ale na jiném povrchu. Tenis se hraje zpravidla na čtyřech typech povrchů (beton, antuka, tráva a koberec) a i grandslamové turnaje se liší typem povrchu. Australian Open se hraje na tvrdém syntetickém povrchu, French Open na antuce, nejznámější grandslamový turnaj Wimbledon se hraje na trávě a poslední US Open („United States

Open Tennis Championships“) se hraje také na tvrdém povrchu stejně jako první zmiňovaný, ale na rozdíl od něj je povrch kurtu akrylátový (International Tennis Federation, 2019).

### **1.1.2 Bodový systém**

Bodový systém tenisové hry má hierarchickou strukturu, díky které je vhodný pro vytváření matematických modelů. Poskytuje dostatek různorodých dat, což usnadňuje historicky posoudit výkony jednotlivých hráčů. Díky hierarchické povaze bodového systému je možné zjednodušit modely pomocí vytváření hierarchických struktur.

Tenisový zápas vyhrává ten, kdo jako první dosáhne vítězství v počtu setů, které jsou dány prestiží turnaje. Set získá hráč, pokud vyhraje šest her, případně v nerozhodném stavu 6:6 zvítězí v tie-breaku (Simmonds & O'Donoghue, 2018).

### **1.1.3 Bodování ve hře**

Bodování tenisové hry je založeno na principu posloupnosti bodů, přičemž skóre začíná na 0:0. Pokud podávající hráč vyhraje hru čistou posloupností, pak je to 15:0, 30:0, 40:0 a následuje vítězný míček. Pokud nastane situace, že hra je vyrovnaná 40:40, přichází takzvaná shoda. Aby hráč danou hru vyhrál, musí vyhrát o dva míčky. V tenisové terminologii je první míček výhoda. Pokud hráč vyhraje první míček, říkáme, že má výhodu. V případě, že další míček prohraje, výhodu ztratí a skóre je opět 40:40.

Za stavu 6:6 na hry v setu nastává tie-break. V tie-breaku se začíná skóre od 0:0 a každý míč se počítá jako jeden bod. Hráč, který jako první získal sedmý bod a zároveň má navíc nejméně dvoubodový náskok vůči soupeři (například 7:3 nebo 13:11). Vítěz tie-breaku celý set v poměru 7:6 (Simmonds & O'Donoghue, 2018).

Skóre zápasu je počtem setů vyhraných kterýmkoli hráčem. Aby hráč zvítězil v zápasu, musí vyhrát dva sety jako první. Existují ale i výjimky v pravidlech různých turnajů. Například v mužském grandslamu musí hráč vyhrát tři sety, aby vyhrál zápas. Tyto zápasy budou předmětem našich výpočtů.

### **1.1.4 Pořadí podání**

Velmi důležitým faktorem ovlivňujícím tenisovou hru je podání, jelikož podávající hráč má velkou výhodu v uvedení míče do hry. Pokud je podání rychlé a silné, uvede to přijímacího hráče rovnou do obranného postavení, které je pro něj méně výhodné. Velkou zvláštností této hry je, že pokud hráč zkaží svůj první pokus, tak ho může

beztrestně opakovat. V tom případě už ale podání většinou není tak silné, aby hráč neudělal dvě chyby za sebou. Jednalo by se totiž o dvojchybu, která znamená ztrátu bodu.

Po skončení první hry se stává přijímající hráč podávajícím a podávající hráč přijímajícím. To se střídavě opakuje ve všech dalších hrách zápasu. V tie-breaku se podání střídá po dvou odehraných míčích (Pollard, 2017).

### **1.1.5 Tenisové turnaje**

Do větších turnajů se hráči dostávají díky počtu bodů, které postupně sbírají na všech turnajích. Existuje několik úrovní tenisových turnajů. V této diplomové práci se zaměříme na druhou nejvyšší úroveň turnajů. První (nejvyšší) úrovní je turnaj mistrů, kde se utkává nejlepších 8 tenistů a koná se jednou ročně vždy na konci sezóny. Druhým největším turnajem jsou grandslamy, které jsou v sezóně čtyři (pokaždé na jiném povrchu) a do hlavního turnaje postupuje 128 hráčů. Další úrovní jsou turnaje ATP Masters, které jsou rozdělené na ATP 1000, ATP 500 a ATP 250 podle toho, kolik bodů si připsá vítěz turnaje.

Všechny profesionální turnaje se hrají pavoukovým systémem. To znamená, že v každém kole vypadne polovina hráčů. V některých případech nemusí být tento systém výhodný, protože hráč může mít „těžký los“ a například hned v prvním kole může narazit na těžkého soupeře, se kterým vypadne. Naopak pokud by ho porazil, v dalších kolech by mohl narazit na jednodušší soupeře, než potkal v prvním kole (Spanias, 2014).

#### **1.1.5.1 Nasazení hráči**

Nasazení hráčů v turnaji je vytvářeno strategicky, a to například z toho důvodu, aby se hráč, který je ve světovém žebříčku na prvním místě, nepotkal s hráčem na druhém místě hned v prvním kole.

To znamená, že hráči z prvního místa světového žebříčku se přidělí první místo v pavoukovi a hráči, který je druhý v žebříčku, se přiřadí poslední místo v pavoukovi. Tímto je zajištěné, že pokud se tito dva hráči potkají, tak až ve finálovém zápase (Grand Slam Board, 2020).

#### **1.1.5.2 Přímí postupující**

Přímo do turnaje postupují hráči, kteří získali dostatečný počet bodů, aby byli na vrcholu žebříčků. Počet přímo postupujících ovlivňuje velikost turnaje, která souvisí s počtem startujících hráčů (Grand Slam Board, 2020).

### **1.1.5.3 Kvalifikace**

Před každým Grand Slamem se hraje kvalifikační turnaj, ve kterém mají šanci dostat se do turnaje i hráči, kteří nejsou v žebříčku mezi top hráči. Kvalifikační turnaj se hraje ve třech kolech a vítězové postupují do hlavní části turnaje. Počet kvalifikovaných hráčů záleží opět na velikosti turnaje (Grand Slam Board, 2020).

### **1.1.5.4 Šťastní poražení**

Občas se stane, že hráč, který se dostal do turnaje jakýmkoliv z výše uvedených způsobů, se nemůže turnaje ze zdravotních nebo jakýchkoliv jiných důvodů zúčastnit. V tom případě na jeho místo nastoupí některý z hráčů kvalifikace, který byl poražen v posledním kole. Proto se pro tohoto hráče užívá termín „šťastný poražený“ (Grand Slam Board, 2020).

### **1.1.5.5 Divoká karta**

Dalším specifickým prvkem tenisu je tzv. divoká karta. Divokou kartu dostávají hráči podle uvážení organizátorů turnaje. Kdo tuto kartu dostane, má možnost se zúčastnit hlavního turnaje bez nutnosti projít kvalifikačním sítím. Divoké karty se obvykle dávají hráčům, kteří si v minulosti vedli dobře, ale klesli v žebříčku. Mohou být také věnovány místním talentům nebo místním oblíbeným hráčům. Počet divokých karet, které se mohou udělit, je zpravidla 3 až 6 (Grand Slam Board, 2020).

## **1.2 Oficiální bodovací systém hráčů**

V každém sportu jsou vítězové a poražení, kteří jsou v tabulkách. V tenise hráči dostávají body podle toho, do jaké fáze turnajového pavouka se dostanou a také v závislosti na prestiži turnaje. Podle počtu bodů pak mohou hráči ale i fanoušci sledovat, který tenista je na tom lépe. Také se podle těchto bodů, a tedy umístění v žebříčku, určují účastníci velkých turnajů, kam je třeba se kvalifikovat.

ATP („Association of Tennis Professionals“) žebříček je oficiální pořadí tenistů, který sestavuje Asociace tenistových profesionálů podle dlouhodobých výsledků turnajů. Započítávání bodů do žebříčků se v historii několikrát měnilo. Poslední aktualizace proběhla před sezónou 2021. Body v žebříčku zůstávají započítány 52 týdnů a připisují se za umístění v 19 turnajích: grandslamech, ATP, Challengerech i nižších úrovních ITF („International Tennis Federation“). Osm nejlepších hráčů, kteří se dostanou na závěrečný Turnaj mistrů, mají místo 19 započítáno 20 turnajů. Aktualizace žebříčku



probíhá každé pondělí. Detailní systém rozdělování bodů však není pro tuhle práci důležitý.

Před zahájením turnaje vedl tento žebříček Novak Djokovič s 12030 body. Druhý byl Rafael Nadal s 9850 body a třetí Thiem Dominic měl 9125 bodů (Utathya, 2021).

### **1.3 Australian Open**

Jedním z nejznámějších a zároveň nejprestižnějších turnajů v tenise je grandslam, ve kterém se potkává absolutní špička profesionálních tenistů. Tento turnaj zahrnuje v jedné sezóně čtyři samostatné turnaje, pokud se nestane nějaká neočekávaná situace. Patří mezi ně Australian Open, French Open, Wimbledon a US Open (Associated Press, 2021; Tennis Australia, 2021a). I na tomto turnaji se hraje vyřazovacím systémem. To znamená, že pokud hráč prohraje, na turnaji končí (Červinka, 2017).

Australian Open 2021 se konal od 8. do 21. února 2021 v australském Melbourne Parku. Hrál se na 16 hřištích s tvrdým povrchem. Turnaj byl zorganizován Mezinárodní tenisovou federací a Australský tenisovým svazem.

Na tomto turnaji byl každý dvorec opatřen 12 kamerami, které nahradily čárové rozhodčí. Tento pokrok byl již dlouho očekáván, ale značně ho urychlila pandemie onemocnění COVID-19. Poprvé v historii grandslamu tak byl využit systém elektronického čárového rozhodčího s nahranými hlasovými projevy, jež byly na dvorci hlášeny v reálném čase. Při dopadu míče v těsné blízkosti čar tak zvané „jestřábí oko“ zobrazilo stopu úderu na velkoplošné obrazovce. Vizualizace se aktivovala po konci výměny při dopadu míče méně než 150 mm od čáry, respektive u podání při dopadu do vzdálenosti 50 mm od čáry. Tenista měl možnost požádat o přezkum.

Díky systémů automatického rozhodčího odpadají situace, ve kterých byl míček vyhodnocen obráceně, než si hráč myslel, a v důsledku opačného výroku rozhodčího byla ovlivněna psychika hráče. Tato nečekaná událost v zápase se dá velmi těžko předpovědět. Pro predikční modely je tedy velmi těžké tuto stránku hráče odhadovat (Associated Press, 2021; Tennis Australia, 2021a).

### **1.4 Finanční odměny v turnaji**

Na grandslemech jsou odměny za postup do dalších kol vyšší než na jiných turnajích. To je jeden z hlavních důvodů, proč byl tento turnaj vybrán. Předpokládá se, že v těchto turnajích budou hráči daleko méně náchylní k podvodům.

Dle dostupných informací (ATP Tour, 2020; peRFect Tennis, 2022) byly odměny pro Australian Open 2021 následující:

	Odměna (USD)	Odměna (Kč)	Rozdíl oproti roku 2020
Vítěz	2 138 125	32 713 312	-33,25 %
Poražený finalista	1 166 250	17 843 625	-27,36 %
Semifinále	660 875	10 111 387	-18,27 %
Čtvrtfinále	408 188	6 245 276	0,00 %
Čtvrté kolo	248 800	3 806 640	6,67 %
Třetí kolo	167 163	2 557 594	19,44 %
Druhé kolo	116 625	1 784 362	17,19 %
První kolo	77 750	1 189 575	11,11 %

Tabulka 1 Tabulka turnajových odměn Australian Open 2021.

## 1.5 Zákon o hazardních hrách

Tento zákon najdeme pod číslem 186/2016 Sb, o hazardních hrách. Upravuje hazardní hry a jejich druhy, podmínky jejich provozování, opatření pro zodpovědné hraní a působnost správních orgánů v oblasti provozování hazardních her.

Hazardní hrou se podle zákona o hazardních hrách rozumí hra, sázka nebo los, do nichž sázející vloží sázku, jejíž návratnost se nezaručuje, a v nichž o výhře nebo prohře rozhoduje zcela nebo zčásti náhoda nebo neznámá okolnost.

Zákon upravuje tyto druhy hazardních her:

- a) loterii,
- b) kursovou sázku,
- c) totalizátorovou hru,
- d) bingo,
- e) technickou hru,
- f) živou hru,
- g) tombolu,
- h) turnaj malého rozsahu.

Pro účely této práce se stačí seznámit s kursovým sázením. Kursová sázka je podle § 28 hazardní hra, u níž je výhra podmíněna uhodnutím sázkové příležitosti. Sázkovou příležitostí se rozumí zejména sportovní výsledek nebo událost veřejné pozornosti. Výše výhry je přímo úměrná výhernímu poměru, ve kterém byla sázka přijata, a výši sázky. Při této sázce nastane pouze jedna správná odpověď.

Také je možné vsázet při sledování přímého přenosu nebo při sledování zápasu přímo na hřišti. Živá kursová sázka je druh kursově sázky, na kterou jsou vklady přijímány v průběhu konání sázkové příležitosti. Sázkový tiket živé kursově sázky musí obsahovat přesný časový údaj o přijetí vkladu na sázkovou příležitost. Provozovatel o každé sázkové příležitosti, na kterou přijal živou kursovou sázku, pořídí datový záznam se současným záznamem přesného času. Provozovatel je povinen uchovávat tento záznam po dobu 1 roku.

## 1.6 Predikce

Predikce výsledků je tu s námi nejméně od starověku. Například v dobách prvních olympijských her lidé odhadovali, kdo se stane vítězem. Moderní technologie posunuly možnosti sázení o velký kus dále. V dnešní době jsou například k dispozici veřejně přístupné webové stránky, které sledují různé statistiky hráčů. Na základě těchto dat vznikají stále propracovanější modely predikce.

Predikce je pojem známý nejen ze statistiky, ale obecně se dá říci, že je to výrok o události, která nastane v budoucnu. Synonymem ke slovu predikce může být například předpověď či prognóza (ABZ.cz, 2021; Petráčková & Kraus, 2000). Předpověď není založena na náhodě (hádání, věštba), ale vychází ze znalosti současnosti a minulosti. Metody předpovědi se liší, ale často vycházejí z vědeckých teorií. Oblast, která se predikcemi hodně zabývá, je statistika. Predikce je ale chápána v širším slova smyslu než klasická předpověď budoucnosti. Statistika konstruuje predikce i v situaci, kdy neznámé hodnoty nesouvisí s budoucím časem. Např. statistika stanoví pravděpodobnost určité diagnózy u pacienta v závislosti na výsledcích fyziologických měření. U konkrétního pacienta pak predikujeme, zda diagnóza nastala či nikoliv. Nejedná se tedy o předpověď budoucnosti, ale i současného neznámého stavu. Na druhou stranu se statistika omezuje pouze na číselné předpovědi (Brom, 2018).

## 1.7 Sázka

Dle občanského zákoníku 89/2012 Sb., v oddílu 3, § 2873 – 2880 je sázka zařazena mezi tzv. odvážné smlouvy. Prospěch z ní totiž závisí na nejisté události.

Sázkou se rozumí smluvní vztah, při kterém se každá ze stran zavazuje vyplatit výhru straně, která daný výsledek dokázala určit. Sázky jsou tedy vzájemná protichůdná tvrzení. Výhrou nemusí být pouze peníze, ale i hmotné produkty nebo závazky vykonat nějakou činnost. Sázkou se alespoň jedna strana zavazuje vůči druhé plnit výhru, pokud se ukáže její tvrzení o nějaké skutečnosti, která stranám není dopředu známá, jako nesprávné, nebo tvrzení druhé strany o této skutečnosti se ukáže jako správné. Je očividné, že pokud jedna ze stran při uzavírání smlouvy o sázce má jasné povědomí o tom, jak se věci skutečně mají (předmětem sázky nemusí být totiž jen budoucí skutečnost, ale i minulá, resp. historická, kterou je možno vypátrat např. v historické literatuře či v encyklopedii), nemůže být taková smlouva platně uzavřena. Obě strany tedy musí mít stejnou výchozí pozici a stejnou naději na výhru i prohru. Nicméně, informace, které zvýší šance jednoho účastníka na výhru (pohybuje se v daném oboru, a proto je schopen zvýšit pravděpodobnost svého úspěchu) jsou obecně přípustné (Trojan, 2021).

Od sázek upravených občanským zákoníkem (např. sázky mezi kamarády či kolegy o výsledku hokejového zápasu) musíme odlišovat sázky, hry a losy, k jejichž provozování je nutno získat povolení správního orgánu za současného dodržení celé řady povinností vyplývajících z poměrně nové právní úpravy v podobě zákona o hazardních hrách. Tento zákon přinesl novou koncepční úpravu hazardních her, včetně dvoustupňového povoloovacího řízení. To znamená, že základní povolení k provozování hazardních her vydává Ministerstvo financí, avšak pravomoc rozhodovat o umístění herního prostoru (herna, kasino) na území obce získaly obecní úřady.

Tímto problémem se zabývá i Jiří Bílý ve své knize (Bílý, 2003). Smlouva o sázce vzniká ujednáním o události oběma stranám neznámé a stanovení ceny pro toho, jehož tvrzení odpovídá výsledku. Sázka je neplatná v případě, pokud jedna ze smluvních stran zná výsledek a zatají ho před druhou stranou. V tomto případě by se sázka považovala za darování, protože jedna ze stran by úmyslně prohrála. Smlouvou o sázce se tradičně rozumí smlouva, kterou se jedna strana zavazuje, že poskytne druhé smluvní straně plnění pro případ, že nastala nebo nastane určitá událost. Charakteristickým znakem pojmu „sázka“ jsou protichůdná tvrzení stran o tom, zda sázková událost existovala, existuje

nebo zda nastane. V případě sázky jde o potvrzení sázkové události. Určitá potíže v oblasti smluv o sázce může nastat při rozlišení pojistné smlouvy a sázky. V tomto případě je potřeba zkoumat účel smlouvy. Při sázce se snaží obě strany obohatit se. Rozlišujeme jednostranné a dvojstranné sázky. Při jednostranné sázce slibuje jedna strana druhé straně sjednané plnění v případě, že z protichůdných tvrzení stran vyplyne, že tvrzení slibující strany je nepravdivé. Při dvoustranné sázce si pak navzájem slibují plnění pro případ, že se tvrzení některé z nich ukáže nepravdivým (Mastela et al., 2011).

V novém občanském zákoníku se můžeme setkat s výslovnou definicí sázky. Podle něj se sázkou alespoň jedna strana zavazuje vůči druhé plnit výhru, ukáže-li se nesprávným její tvrzení o skutečnosti neznámé nebo ukáže-li se tvrzení druhé strany o této události správným. Pokud má strana, jistotu o výsledku a zatají to straně druhé, je sázka neplatná. Tato definice se tolik neliší od dosavadní úpravy. Ta stejně, jako ta dnešní, vychází z toho, že skutečnosti jsou stranám neznámé, bez ohledu na to, zda šlo o skutečnost již nastalou nebo mající teprve nastat v budoucnu. Jeden z paragrafů představuje určitý návrat ke kořenům (Vlček, 2015).

## **1.8 Výhodnost sázky**

Pokud se sází na dva vyrovnané hráče a na jednoho je kurz například 1,08, je jasné, že riskují daleko více, než mohou vyhrát. (BetArena.cz, 2015)

Pro lepší pochopení počítání výhodnosti sázky může sloužit následující příklad: Budeme házet mincí. Jaká je férová hodnota kurzu, že padne určená strana? Víme, že máme padesátiprocentní šanci neboli  $\frac{1}{2}$ . Po přepočtu na decimální kurz nám vyjde 2,00. Pokud by nám sázková kancelář nabídla jakýkoliv vyšší kurz, měli bychom sázku přijmout, protože se jedná o sázku s výhodným kurzem. Kdyby však například sázková kancelář nastavila kurz 2,20, znamenalo by to, že je podle ní šance na výhru 45,45 procent. My ale víme, že šance je 50 procent, takže jsme našli value. Jakýkoliv kurz, nižší než dva, se naopak nevyplatí.

## **1.9 Převod kurzů na pravděpodobnosti**

Tabulka v této kapitole ukazuje posloupnost kurzů a výše pravděpodobností výhry. Přepočet kurzů na pravděpodobnosti se počítal v intervalu 1 až 1,85. Kurz 1 znamená, že pravděpodobnost vítězství tohoto hráče je 100 % a kurz 1,85 je naprosto vyrovnané utkání a oba hráči mají shodnou pravděpodobnost, tedy 50 %. Určení pravděpodobností

nám postačí pouze pro kurzy v tomto intervalu, protože probabilita výhry hráče, na kterou je kurz vyšší, se spočítal odečtením pravděpodobnosti výhry favorita utkání od čísla 1.

Kurz	=	Pravděpodobnost	Kurz	=	Pravděpodobnost	Kurz	=	Pravděpodobnost
1	=	100 %	1,29	=	83 %	1,58	=	66 %
1,01	=	99 %	1,3	=	82 %	1,59	=	65 %
1,02	=	99 %	1,31	=	82 %	1,6	=	65 %
1,03	=	98 %	1,32	=	81 %	1,61	=	64 %
1,04	=	98 %	1,33	=	81 %	1,62	=	64 %
1,05	=	97 %	1,34	=	80 %	1,63	=	63 %
1,06	=	96 %	1,35	=	79 %	1,64	=	62 %
1,07	=	96 %	1,36	=	79 %	1,65	=	62 %
1,08	=	95 %	1,37	=	78 %	1,66	=	61 %
1,09	=	95 %	1,38	=	78 %	1,67	=	61 %
1,1	=	94 %	1,39	=	77 %	1,68	=	60 %
1,11	=	94 %	1,4	=	76 %	1,69	=	59 %
1,12	=	93 %	1,41	=	76 %	1,7	=	59 %
1,13	=	92 %	1,42	=	75 %	1,71	=	58 %
1,14	=	92 %	1,43	=	75 %	1,72	=	58 %
1,15	=	91 %	1,44	=	74 %	1,73	=	57 %
1,16	=	91 %	1,45	=	74 %	1,74	=	56 %
1,17	=	90 %	1,46	=	73 %	1,75	=	56 %
1,18	=	89 %	1,47	=	72 %	1,76	=	55 %
1,19	=	89 %	1,48	=	72 %	1,77	=	55 %
1,2	=	88 %	1,49	=	71 %	1,78	=	54 %
1,21	=	88 %	1,5	=	71 %	1,79	=	54 %
1,22	=	87 %	1,51	=	70 %	1,8	=	53 %
1,23	=	86 %	1,52	=	69 %	1,81	=	52 %
1,24	=	86 %	1,53	=	69 %	1,82	=	52 %
1,25	=	85 %	1,54	=	68 %	1,83	=	51 %
1,26	=	85 %	1,55	=	68 %	1,84	=	51 %
1,27	=	84 %	1,56	=	67 %	1,85	=	50 %
1,28	=	84 %	1,57	=	66 %		=	

Tabulka 2 Hodnota kurzů vzhledem k pravděpodobnosti

## 1.10 Strategie sázení

Sázení je pro mnoho lidí nejen koníčkem a zábavou, ale také formou zajímavého přivýdělku. Existují různé druhy strategií sázení a každá z nich se řídí pevně nastavenými pravidly.

Při volbě strategie sázení by měl sázející zvážit své zkušenosti, znalosti i finanční možnosti. Vhodně zvolená strategie mu poté z dlouhodobého hlediska přinese zisk. Oproti tomu chybně zvolená strategie sázejícího připraví o spoustu peněz.

V následujících kapitolách jsou blíže popsány nejznámější a nejvyužívanější strategie sázení.

### **1.10.1 Strategie stejný vklad**

Strategie stejný vklad patří mezi nejjednodušší strategii sázení. Jak už její název napovídá, je založena na principu, že pokaždé sázející vsadí stejnou částku. Její velkou nevýhodou je to, že při výpočtu kurzové sázky není zohledněna výše kurzu. To znamená, že při kurzu 1,1 by sázející vsadil zcela stejnou částku jako např. při kurzu 6,5.

U této strategie se doporučuje vsadit maximálně 0,9 % z celkové výše banku. Pokud bank je 100 000 pak v tomto případě je to 900,- s jakýmkoliv kurzem.

Díky této strategii sázející v případě prohry nepřijde o mnoho peněz v porovnání s ostatními strategiemi sázení. Zároveň ale neposkytuje možnost, jak vyhrát velký obnos peněz (Sazkovekancelare.cz, 2020).

### **1.10.2 Strategie vsadit vše – all in**

Strategie vsadit všechno se řadí mezi velmi riskantní strategie, která ale v případě jistoty výhry může vést k rychlému a velkému zisku. Tato strategie je založena na tom, že sázející vsadí veškerou finanční hotovost, kterou má vyčleněnou na sázení, a poté čeká na výsledek. V případě výhry ho čeká velký zisk, ale pokud prohraje, přijde o všechno.

### **1.10.3 Strategie Martingale**

Principem této strategie je, že při každé prohře sázející zdvojnásobí sázku. Při výhře se vše resetuje a začíná se od znova. Jestliže sázející vyhraje, vyrovná tím předchozí ztráty, a navíc může vyhrát pěkný obnos peněz. Nicméně pokud sázející nebude mít štěstí, jeho prohry se budou zvyšovat.

Pokud tedy sázející zvolí jako vstupní sázku 100,-, při prohře se každá další sázka zvýší dvakrát, tedy bude 200,-; 400,-; 800,- atd. Jestliže u tiketu za 800,- vyhraje, zisk bude 100,- a ztráta 0,-. Pak sázející začíná od začátku. Z toho vyplývá, že čím vyšší bude první sázka, tím vyšší bude zisk.

Při vhodném nastavení rozpočtu je tedy strategie Martingale účinný způsob, jak získat hodně peněz. Je ale důležité vsadit tiket s minimálním kurzem 2,0 a vyšším (Sazkovekancelare.cz, 2020).

#### 1.10.4 Strategie Fibonacci

Fibonacciho posloupnost byla popsána italským matematikem Leonardem Fibonaccim již ve 13. století. Tato posloupnost je postupně zvyšující se řadou, která začíná čísly 0 a 1 a ve které je každé číslo rovné součtu dvou předchozích čísel, např. 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377 (Mitchell, 2022).

Strategie sázení Fibonacci je podobná předešlé Martingale strategii. Kurzové sázky se řídí výše zmíněnou Fibonacciho číselnou řadou při každé prohře. Kurz sázení by měl v tomto případě být minimálně 2,0, ideálně vyšší. To znamená, že pokud sázející vsadí tiket v hodnotě 100,-, tak při prohře navýší sázku dle Fibonacciho posloupnosti na 200,-. V případě další prohry dále vsadí 300,- atd. Tato strategie může vést k poměrně vysokému zisku, pokud bude mít správně nastavený rozpočet a bude výhodný kurz (Pinnacle, 2013; Sazkovekancelare.cz, 2020).

#### 1.11 Data získaná ze statistik

Díky zdokonalování technologií používaných v zápasech se množství dohledatelných statistik rozšiřuje, a navíc jsou aktuální ihned po utkání. Na oficiálních webech těchto turnajů se dají dohledat tyto statistiky:

Na podání hráče A

- Počet es
- Počet dvojchyb
- Úspěšnost 1. podání včetně esa
- Body po 1. podání včetně esa
- Body po 2. podání
- Odvrácené break bally

Na příjmu hráč A

- Body po 1. podání soupeře
- Body po 2. podání soupeře
- Využité break bally

Body hráče A

- Vítězné míče
- Nevynucené chyby

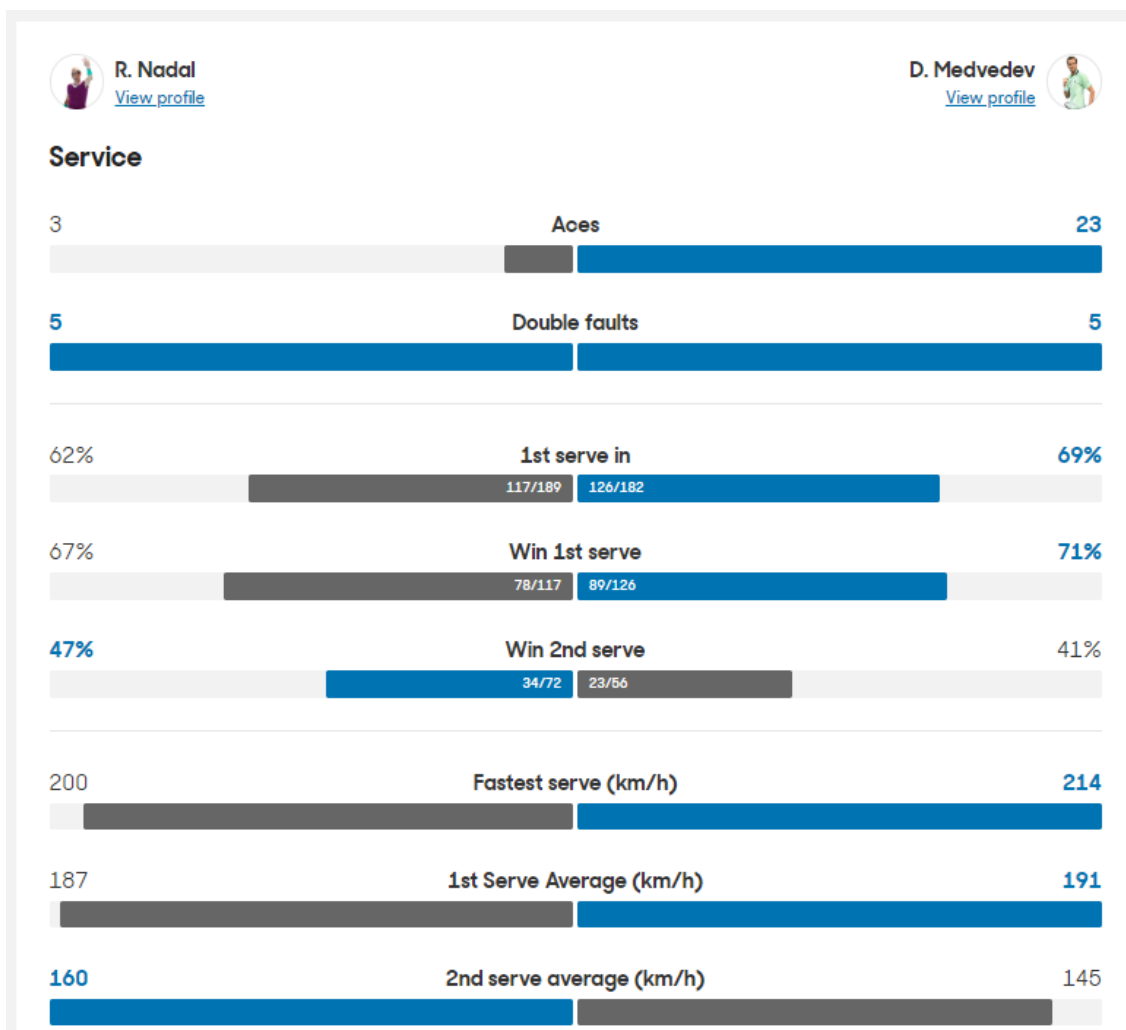


- Vyhrané míče na síti
- Nejvíce míčů v řadě
- Vyhrané míče na podání bez es
- Vyhrané míče na příjmu
- Vyhrané míče celkem

#### Gamy hráče A

- Nejvíce gamů v řadě
- Vyhraná podání
- Vyhrané gamy na příjmu
- Vyhrané gamy celkem
- Všechna tato data se ukládala pro případné vylepšení modelu. Abych vypočítal tento model, potřeboval jsem pouze tato data:
  - Esa a esa soupeře
  - Dvojchyby a dvojchyby soupeře
  - Úspěšnost 1. podání včetně esa
  - Body po 1. podání včetně esa
  - Body po 2. podání
  - Body po 1. podání soupeře
  - Body po 2. podání soupeře
  - Celkový počet míčků

Na obrázku 2 je vidět originální tabulka s daty turnaje. Data jsou přehledně poskládaná i s ukazateli procentuálních hodnot. Všechna tato data jsou k dispozici ihned po utkání na oficiálních stránkách turnaje Australian Open ([ausopen.com](http://ausopen.com)).



Obrázek 2 Hodnoty zobrazené na oficiálních stránkách soutěže: <https://ausopen.com/>.

V následující tabulce Tabulka 3 Statistika prvních 4 kol hráče: Dimitrov Grigor jsou vidět všechna data účastníků čtvrtfinále společně s pomocnými výpočty před čtvrtfinálovým zápasem. Tabulka ukazuje data Dimitrova Grigora, který se stal jedním z čtvrtfinalistů tohoto turnaje. Je z ní patrné, že třetí kolo odehrál pouze jeden set a poté soupeř ze zápasu odstoupil.

Jméno	Dimitrov Grigor	Žebříček	19	29 let	06.05.1991	81kg	191cm	Bulharsko
Soupeř	Čilič Marin	Bolt Alex	Carreno Busta Pablo	Thiem Dominik				
Žebříček	43	187	16	3				
Výsledek	3:0	3:0	1:0 Nedohráno	3:0				
Počet gamů	30	27	6	26				
Kolo	1	2	3	4				
Počet míčků	184	172	48	179				
<b>Podání</b>	<b>Počet</b>	<b>Hodnoty</b>	<b>Počet</b>	<b>Hodnoty</b>	<b>Počet</b>	<b>Hodnoty</b>	<b>Výpočty</b>	
Esa	19		10		3		Pav 145,75	
Dvojitky	2		4		0		z 10,50	
Úspěšnost 1. podání včetně esa	69%		68%		56%		y 1,75	
Body po 1. podání včetně esa	83%	52 z 63	81%	47 z 58	90%	9 z 10	x 3,25	
Body po 2. podání	54%	15 z 28	63%	17 z 27	50%	4 z 8	w 1,75	
Odvraćení break bally	75%	3 z 4	100%	4 z 4	100%	1 z 1	a 0,07	
<b>Příjem</b>							b 65%	
Body po 1. podání soupeře	22%	12 z 54	45%	27 z 60	63%	15 z 24	c 81%	
Body po 2. podání soupeře	49%	19 z 39	37%	10 z 27	67%	4 z 6	d 0,0120	
Využití break bally	40%	4 z 10	44%	4 z 9	50%	4 z 8	e 56%	
<b>Body</b>							f 0,0223	
Vítězné míče	34		26		10		g 70%	
Nevyručené chyby	19		18		6		h 43,75%	
Vyhrané míče na síti	100%	4 z 4	89%	8 z 9	60%	3 z 5	i 53,50%	
Nejvíce míčů v řadě	9		13		7		j 0,012	
Vyhrané míče na podání bez es	74%	67 z 91	75%	64 z 85	72%	13 z 18		
Vyhrané míče na příjmu	33%	31 z 93	46%	40 z 87	63%	19 z 30		
Vyhrané míče celkem	54%	100 z 184	60%	104 z 172	67%	32 z 48		
<b>Gamy</b>								
Nejvíce gamů v řadě	3		7		7			
Vyhraná podání	93%	14 z 15	100%	14 z 14	100%	3 z 3		
Vyhrané gamy na příjmu	27%	4 z 15	31%	4 z 13	100%	4 z 4		
Vyhrané gamy celkem	60%	18 z 30	67%	18 z 27	100%	7 z 7		

Tabulka 3 Statistika prvních 4 kol hráče: Dimitrov Grigor

V prvním řádku tabulky můžeme vidět zjištěné hodnoty hráče, kterého se daná tabulka týká. Žebříček ATP ukazuje dlouhodobou formu hráče, protože body do tohoto žebříčku se sbírají celoročně a nenulují se po skončené sezóně. Nereflektuje však aktuální formu hráče. Například pokud mám hráče na dvacáté pozici, nevím, zda se na tuto pozici dostal z první desítky, a tedy klesá žebříčkem a nemá aktuální formu, nebo se posunul z padesátého místa.

Hodnoty váha, věk a výška jsou hodnoty, které by se mohly využít v další fázi. Obzvláště výška hráče má velký vliv na utkání, protože hráč s vyšší výškou dokáže odehrát podání daleko nebezpečněji. Národnost hráče může hrát velkou roli v případě, že se turnaj hraje na domácí půdě, jako bylo například vystoupení domácího bouřliváka Nika Kyrgiose, který potrápil, i díky výborné atmosféře, ve třetím kole i Dominika Thiema v pětisetové bitvě.

V další části tabulky jsou údaje o soupeři a výsledek daného kola, případně informace, že se utkání nedohrálo. Zde jsou ale také spíše informativní čísla, která se nepoužívají do našich výpočtů.

V poslední části jsou uvedena data z daného utkání. Buňky zvýrazněné modrou barvou jsou data, která jsou ve výpočtech predikčního modelu. Bílá data jsou zde uvedena hlavně z důvodu zkoušení úpravy modelu, kde jsem si zkoušel do výpočtů přidávat různé hodnoty, vzorce atp. Bylo tedy jednodušší mít všechna data v tabulce, než pokaždé zpětně dohledávat v každém utkání potřebnou statistiku. V závěru této kapitoly přikládám predikci dalších tenisových turnajů, ve kterých už je tato tabulka přizpůsobena tomuto modelu a zadávají se do ní pouze potřebné hodnoty, které se skutečně používají ve výpočtech.

V tabulce je statistika prvních 4 kol před čtvrtfinále a pomocné výpočty, se kterými se dále pracuje v kapitole Statistiky zápasu. Výpočet těchto hodnot je popsán v kapitole Low-level point model.

Tyto statistiky ukazují také na stabilitu hráčů. Pokud je hráč schopný dávat v každém zápase velké množství es, může si tím velmi pomoci bez vynaložení větších sil. Proti této statistice jde ukazatel dvojchyb. Pokud hráč dá do prvního podání veškerou sílu a nepovede se mu, druhé už nesmí zkazit, protože bod získá soupeř, a tak musí podat přesněji, což v naprosté většině případů znamená současně i pomaleji. To znamená, že soupeř bude mít lepší pozici při vstupu do hry, protože na něj poletí pomalejší míček.

Jak si první hráč s tímto případem poradí ukazuje řádek body po druhém podání. Pokud hráč neumí podávat kvalitně, musí si umět poradit ve hře.

Proti tomu stojí statistiky tohoto hráče na příjmu, kde je vidět, jak si umí poradit s prvním a druhým podáním soupeře.

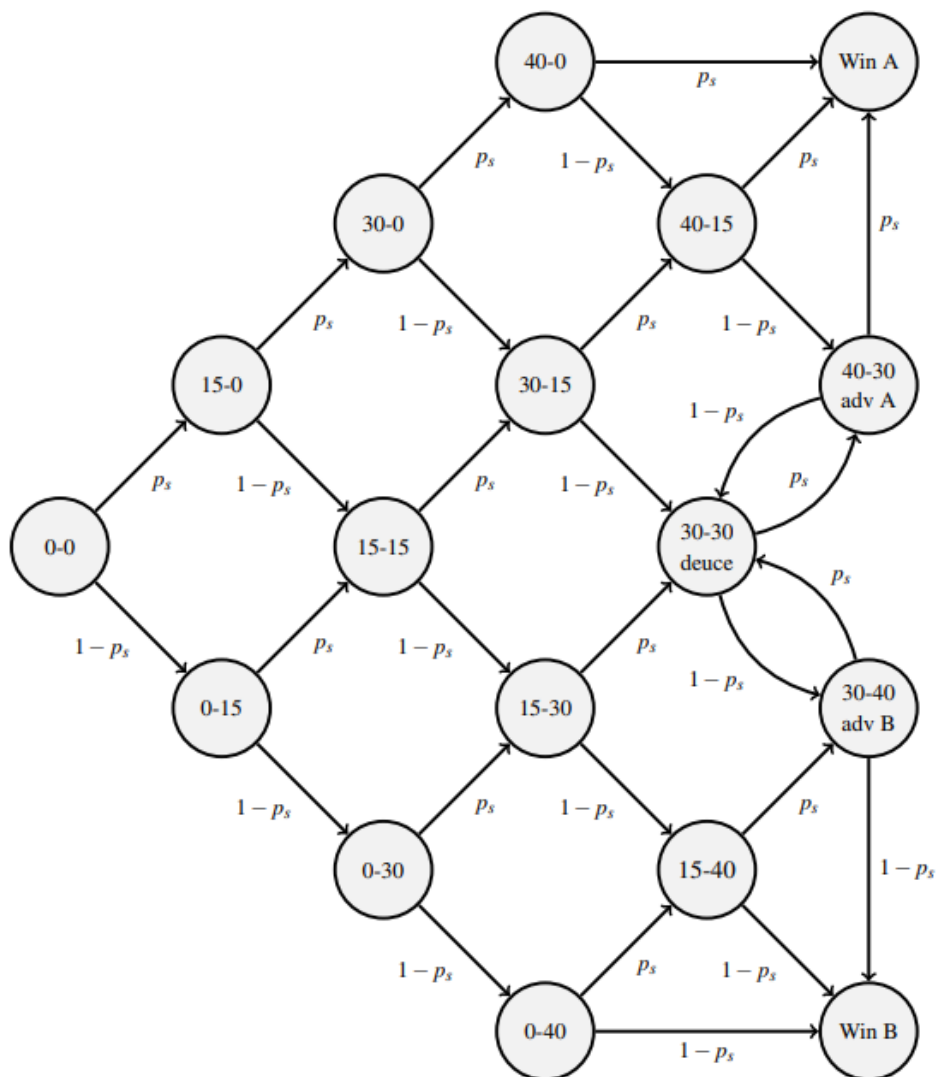
Buňky, které jsou v tabulce zvýrazněny modře obsahují hodnotu, se kterou se počítá v následujících výpočtech pro predikci.

Další tabulky čtvrtfinalistů Australian Open 2021 jsou v přílohách číslo 1 – 7.

## **1.12 Markovovy řetězce**

Markovovy řetězce představují speciální typ náhodných procesů, které slouží pro popis chování neočekávaných jevů probíhajících v čase. V případě Markovových řetězců se pak jedná o náhodný proces bez paměti: stav systému v budoucnu závisí pouze na současném stavu a nikoli na minulosti systému. Tato vlastnost má některé důležité důsledky, mimo jiné umožňuje za určitých podmínek určit rozdělení pravděpodobností stavů při ustálení řetězce, tzv. stacionární rozdělení (Fačevicová et al., 2018).

Predikce výsledku tenisového turnaje s použitím Markovových řetězců je založena na principu, v němž stav odpovídá skóre a pravděpodobnosti přechodu oproti pravděpodobnosti zisku nebo ztráty bodu. Potom lze odvodit uzavřené rovnice, které poskytují způsob, jak převést pravděpodobnosti výsledku bodu, hry nebo setu na pravděpodobnosti výsledku zápasu. Tento princip byl využit např. v práci autorů Newtona a Kellera a je využit i v této práci. Následující obrázek č. 2 znázorňuje Markovův řetězec pro tenisovou hru. (Newton & Keller, 2005).



Obrázek 3 Markovův řetězec modelující tenisovou hru. Převzato z (Sipko & Knottenbelt, 2015).

### 1.13 Metody predikce tenisových utkání

V těchto kapitolách jsou popsány metody predikce vítězů. Jsou zde popsány metody jednoduché, ale i složité, ke kterým jsou potřeba větší znalosti statistiky a matematiky.

#### 1.13.1 Win / Loss index

Dle tohoto indexu jsou hráči sledováni v posledních 52 týdnech a porovnává se počet pohárů, který hráč získal, počet výher, proher a z toho vyplývající poměr (Clarke & Dyte, 2000). Tato statistika by se dala využít při porovnávání formy hráčů. Nicméně, tento model je nepraktický. Do zápasu vstupuje spousta dalších vlivů například prestiž turnaje, povrch a další. Speciálně v roce 2021 je statistika za poslední rok nepříjemně zkreslená.

### 1.13.2 Head to head analýza

Head to head analýza slouží k jednoduchému porovnání dvou hráčů hrajících proti sobě. Jsou v ní zohledněna data z předchozích her, styl hry hráčů, umístění hráčů, aktuálním umístěním v ATP a umístěním ve WTA (Tennis Board, 2022).

K této analýze je dostupné velké množství aktuálních statistik (Matchstat.com, 2022).

### 1.13.3 Magnus-Klaassen model

Autoři Magnus a Klaassen (Klaassen & Magnus, 1998) představili „logitový model“. Tento model byl vyvinut s předpokladem náhodných veličin. Pro představení logitového modelu se označí oba konkurenční hráči jako  $A$  a  $B$  a  $P_{AB}$  jako pravděpodobnost, že  $A$  vyhraje bod proti  $B$  (přitom může platit  $P_{AB} \neq P_{BA}$ ). Předpokládá se, že  $P_{AB}$  a  $P_{BA}$  zůstanou fixní během zápasu, díky čemuž můžeme vypočítat pravděpodobnost, že  $A$  vyhraje zápas proti  $B$ . Cílem je odhadnout tyto pravděpodobnosti.

Nejprve určíme proměnnou  $RANK_i$ . Ta označuje hodnocení hráče  $i$ , které bylo publikováno v ATP těsně před zápasem. ATP zveřejňuje každý týden žebříček nejlepších 200 hráčů, který je založen na jejich výkonu za posledních 52 týdnů. Pak proměnná

$$\bar{R}_i = 8 - \log_2(RANK_i)$$

se bude interpretovat jako „očekávané kolo“ hráče  $i$  v daném turnaji. Například hráč  $i$  s  $RANK_i = 8$  dosáhne pátého kola a prohraje. Naopak u nejvýše nasazeného hráče (tedy  $RANK_i = 1$ ) se očekává, že turnaj vyhraje. Hlavní výhoda tohoto opatření spočívá v tom, že bere v úvahu, že malý rozdíl v hodnocení mezi dvěma hráči na spodním místě žebříčku je mnohem méně významný než stejný rozdíl v hodnocení mezi dvěma vysoce hodnocenými hráči. Díky tomu výpočet lépe koresponduje s realitou. Hodnota  $\bar{R}_i$  může být i negativní, ale to nezpůsobuje žádné problémy. Dále byla zavedena hodnota  $R$ , která označuje kolo v turnaji uvažovaného zápasu, a tak kvalitu hráče je možné definovat následovně:

$$Q_i = \bar{R}_i + \delta \max(R - \bar{R}_i, 0)$$

Kvalita hráče  $i$  se rovná očekávanému kolu plus korekce, která měří změnu kvality v případě, že hráč dosáhl vyššího kola, než by se dalo očekávat podle jeho pořadí v ATP. Parametr  $\delta$  je třeba odhadnout. Samozřejmě se očekává, že bude pozitivní. Dále, pokud  $A$  stojí proti  $B$ , musíme definovat kvalitu  $A$  proti  $B$  pomocí individuální kvality každého hráče:

$$Q_{AB} = \alpha_0 + \alpha_1(Q_A - Q_B) + \alpha_2(Q_A)$$

Jak naznačuje tato rovnice, kvalita hráče  $A$  při podání proti hráči  $B$  nezávisí pouze na relativní kvalitě ( $Q_A - Q_B$ ), ale také na absolutní kvalitě hráče, který podává. Nyní, za předpokladu, že  $P_{AB}$  závisí pouze na  $Q_{AB}$ , můžeme napsat:  $P_{AB} = \Lambda(Q_{AB})$ , kde  $\Lambda$  je monotónně rostoucí funkce,  $\Lambda: R \rightarrow (0,1)$ . Pro náš případ autoři vybrali  $\Lambda$  jako funkci logistické distribuce definovanou jako:

$$\Lambda(x) = \frac{\exp(x)}{\exp(x) + 1}$$

Pro odhad parametrů  $\alpha_0$ ,  $\alpha_1$ ,  $\alpha_2$  a  $\delta$  autoři použili data ze 481 zápasů, které byly odehrány v mistrovských zápasech mužů a žen ve dvouhře ve Wimbledonu v letech 1992 až 1995. Vyvinuli dva samostatné modely, jeden pro muže a jeden pro ženy. Pro muže nakonec získali odhad  $P_{AB}$ :

$$\hat{P}_{AB} = \Lambda(0.4913 + 0.0387(\hat{Q}_A - \hat{Q}_B) + 0.0372\hat{Q}_A)$$

Proměnné  $\hat{Q}_i$  lze získat nahrazením odhadované hodnoty  $Q_i$  při dosazení čísla 0,7684 za parametr  $\delta$ .

#### 1.13.4 Barnett-Clarke model

ATP každý týden publikuje index spolehlivosti FedEx ATP každého hráče. Jedná se o nové statistické měřítko toho, jak hráči hrají na různých površích, turnajích a za různých podmínek, a to jak za posledních 52 týdnů, tak v průběhu celé kariéry. Pro každého hráče existuje tolik ukazatelů, že je možné zjistit, jak tento hráč hraje za téměř každé situace. ATP bohužel nezveřejňuje, jak byly tyto indexy vypočteny pro další statistickou analýzu. Barnett a Clarke (Barnett & Clarke, 2005) použili ke spojení individuálních schopností hráče následující označení (pro hráče  $i$ ):

- $f_i$  je procento bodů, které hráč získal za podání podle posledních 52 týdnů,
- $g_i$  je procento bodů, které hráč získal při returnu podle posledních 52 týdnů.

V zápase s uvažovanými hráči  $i$  a  $j$  použijeme následující označení:

- $f_{ij}$  označuje očekávané procento získaných bodů, když hráč  $i$  podává proti  $j$ ,
- $g_{ij}$  označuje očekávané procento získaných bodů, když hráč  $i$  returnuje od hráče  $j$ .

Vždy musí platit následující rovnice:  $f_{ij} + g_{ij} = 1$ .

Kdykoli je turnaj analyzován, použijeme následující vzorce:



- $f_t$  označuje procento bodů, které všichni hráči vyhráli při podání v turnaji  $t$  předchozí rok,
- $g_t$  vyjadřuje procento bodů, které všichni hráči získali na returnu v předchozím roce na turnaji  $t$ .

Opět musí platit rovnice:  $f_t + g_t = 1$ .

- $f_{av}$  označuje procento získaných bodů, které vyhráli hráči při svém podání v uplynulém roce.
- $g_{av}$  označuje procento získaných bodů, které vyhráli hráči na returnu v uplynulém roce.

Pravděpodobnosti výhry bodu spočítáme následovně:

$$f_{ij} = f_t + (f_i - f_{av}) - (g_j - g_{av})$$

$$g_{ji} = g_t + (g_j - g_{av}) - (f_i - f_{av})$$

Rovnice  $f_{ij} + g_{ij} = 1$  platí.

Jednoduše řečeno, lze očekávat, že průměrné skóre hráče  $i$ , který podává proti hráči  $j$ , se bude rovnat průměrnému počtu bodů všech hráčů turnaje  $t$  v loňském roce (je brán v potaz povrch hřiště) plus rozdíl, kterým průměrné podání hráče převyšuje celkový průměr (zahrnuje i schopnost podání).

### 1.13.5 Baseline Point Model

Velmi podobnou metodou, jako je model v předchozí kapitole, je i model výchozího bodu, který vytvořili stejní autoři – Barnett a Clarke (2005). Ten se využívá jako výchozí bod pro porovnání výkonu novějších modelů. Předpovídá pravděpodobnost vítězství v zápasu převedením pravděpodobností podání pomocí Markovova řetězce popsaného v kapitole 1.12. Autoři vytvořili následující rovnice pro předpověď pravděpodobností podání, které jsou založeny na kombinaci jednoduchých průměrů z historických dat:

$$f_{ij} = f_t + (f_i - f_{av}) - (g_j - g_{av})$$

$$f_i = a_i * b_i + (1 - a_i)c_i$$

$$g_i = a_{av} * d_i + (1 - a_{av})e_i$$

kde  $f_{ij}$  je pravděpodobnost, že hráč  $i$  získá bod při podání proti hráči  $j$  a:

$a_i$  – je procento úspěšných prvních podání pro hráče  $i$ ,  
 $b_i$  – je procento bodů získaných při prvním podání hráče  $i$ ,  
 $c_i$  – je procento bodů získaných při druhém podání pro hráče  $i$ ,  
 $d_i$  – je procento bodů získaných při returnu prvního podání pro hráče  $i$ ,  
 $e_i$  – je procento bodů získaných při returnu druhého podání pro hráče  $i$ ,  
 $f_i$  – je procento bodů získaných při podání pro hráče  $i$ ,  
 $g_i$  – je procento získaných bodů na returnu pro hráče  $i$ ,  
 $a_{av}$  – je průměrné procento prvního podání u všech hráčů,  
 $f_{av}$  – je průměrné procento bodů získaných při podání u všech hráčů,  
 $g_{av}$  – je průměrné procento bodů získaných při returnu všech hráčů,  
 $f_{av}$  – je průměrné procento turnajových bodů získaných při podání.

Všechny tyto parametry ( $a$ – $g$ ) jsou stanoveny na základě pozorovaných výsledků z předchozích zápasů. Pro své výpočty autoři Barnett a Clarke používají údaje poskytované přímo ATP, které jsou založeny na datech ze 70 nejnovějších předchozích zápasů.

### 1.13.6 Corral-Prieto-Rodríguez model

Jak už bylo řečeno, hodnocení hráče může být velmi důležitým faktorem, které ovlivňuje vytvoření predikčního modelu. Toto tvrzení dokázali autoři Julio del Corral a Juan Prieto-Rodríguez (Corral & Prieto-Rodríguez, 2010). V jejich výzkumu autoři rozdělili své proměnné do tří skupin: minulý výkon hráče, fyzické vlastnosti hráče a vlastnosti zápasu. Následně odhadli tři alternativní modely pro muže a ženy samostatně pomocí údajů tenisového utkání Grand Slamu od roku 2005 do roku 2008. Všechny výsledky ukázaly, že modely využívající minulý výkon hráčů předčí ty, které tak neučinily.

Proměnné, které byly použity pro model, byly:

- ROZDĚLENÍ: rozdíl mezi přirozenými logaritmy žebříčku hráče s nižším hodnocením a hráče s vyšším hodnocením (tj.  $\log$  (s nižším hodnocením) hodnocení hráče) -  $\log$  (hodnocení hráče s vyšším hodnocením)

- EXTOP10H / EXTOP10L: jsou fiktivní proměnné s hodnotou jedna, pokud hráč s vyšším / nižším hodnocením byl v některých bodech nad deseti nejlepšími hráči posledních pět let.
- DIFROTOUR: je rozdíl mezi koly dosaženými hráči s vyšším a nižším hodnocením za předchozí rok na stejném turnaji

Tělesné charakteristiky:

- DIFHEIGHT / DIFHEIGHT2: je výškový rozdíl / čtverec výškového rozdílu mezi hráčem s vyšším a nižším hodnocením (v metrech).
- DIFAGE / DIFAGE2: je věkový rozdíl / čtverec věkového rozdílu mezi hráčem s vyšším a nižším hodnocením.
- LEFTL / LEFTH: když hráč s nižším nebo vyšším hodnocením je levák a hráč s vyšším nebo nižším hodnocením je pravák
- BOTHLEFT: když oba hráči hrají levou rukou
- BOTHRIGHT: když jsou oba hráči praváci

Ačkoli tento model obsahuje dostatečný počet proměnných, neznamená to, že si vede lépe než ostatní. Tento model je jediný, který obsahuje fyzické vlastnosti hráčů, a proto se očekávalo, že bude schopen odhalit některé zvláštnosti, které se mohou objevit v některých zápasech, které ostatní modely nedokázaly. Navzdory skutečnosti, že berou v úvahu mnoho faktorů, se zdá, že rozdíl v žebříčku mezi dvěma hráči ovlivňuje výsledek nejvíce. Dále se ukázalo, že tento algoritmus je přesnější, pokud se na hodnocení hráčů použije logaritmus.

### 1.13.7 Point model

Pravděpodobnost, že hráči při svém podání získají body, je základním parametrem v mnoha modelech tenisových zápasů. Není však jednoduché ji vypočítat.

Řešením může být například vytvoření Markovových řetězců popisujících tenisový zápas na úrovni jednotlivých bodů. Jednotlivé metody se liší tvarem Markovových řetězců a způsobem jakým počítají pravděpodobnosti v řetězcích. Například pravděpodobnost esa při podání, pravděpodobnost vstupu do rally a její vítězství a různých dalších. Níže jsou rozebrány metody založené na vytvoření Markovových řetězců a je zde popsáno, jak se pravděpodobnosti počítají pomocí veřejně dostupných statistických údajů

z předchozích zápasů a jejich úpravy pro konkrétní hráče (Spanias, 2014; Spanias & Knottenbelt, 2013).

### 1.13.8 Vanilla Bradley-Terry model

Další model, který bych rád zmínil, je standardní Bradleyho-Terryho model. Tento model popsali autoři McHale a Morton v (McHale & Morton, 2011). Bradley-Terryho model se používá k odhadování výsledku celého zápasu, ale může se rozšířit i na předpověď výsledků na úrovni jednotlivých bodů (Bradley & Terry, 1952). McHale a Morton použili pro určení svých parametrů maximální věrohodnosti. Model je založen na vztahu:

$$P = (r_{ij} = 1 | s_i, s_j) = \frac{s_i}{s_i + s_j}$$

Kde  $r_{ij} = 1$  značí výhru hráče  $i$  proti hráči  $j$  a  $s_i$  a  $s_j$  jsou dovednosti hráče, které jsou parametry modelu. Rovnici lze upravit následujícím způsobem tak, aby dovednosti dosahovaly pouze kladných hodnot.:

$$P (r_{ij} = 1 | s_i, s_j) = \frac{e^{s_i}}{e^{s_i} + e^{s_j}} = \frac{1}{1 + e^{-(s_i - s_j)}} = \sigma (s_i - s_j)$$

kde  $\sigma$  je logistická sigmoidní funkce (označuje se tak podle tvaru funkce)

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Pro data skládající se ze souboru výsledků  $D = \{r_1, \dots, r_n\}$  pro skupinu hráčů  $M = \{1, \dots, m\}$  se schopnostmi  $S = \{s_1, \dots, s_m\}$  může být pravděpodobnost modelu vyjádřena jako:

$$P (D|S) = \prod_{k=1}^n \sigma(s_w^k - s_l^k)$$

kde  $s_w^k$  a  $s_l^k$  jsou příslušné dovednosti vyhrávajícího a prohrávajícího hráče z výsledků  $D$ . Pomocí této rovnice je tedy možné předpovědět danou sadu dovedností hráče (McHale & Morton, 2011).

### 1.13.9 Free Parameter Point model

Ačkoli výše popsáný Bradleyho-Terryho model lze použít na výsledky na úrovni bodů, obecně nerozlišuje, zda je bod získaný na podání nebo příjmu. Free Parameter Point model rozšiřuje získané skutečnosti o předpovědi odlišných pravděpodobností podání a přijímání pro každého hráče. To znamená, že body jsou rozděleny na dvě skupiny; první pro body, kdy hráč podává, a druhou pro body, kdy hráč přijímá.

V tomto modelu je každý hráč reprezentován dvěma parametry. První z nich představuje jeho útočnou dovednost, a druhý znázorňuje jeho obrannou dovednost. Parametry modelu jsou tedy sada útočných schopností  $S = \{s_1, \dots, s_m\}$  a sada obranných dovedností  $B = \{b_1, \dots, b_n\}$ . V tenise lze podání a příjem podání považovat za útočnou a obrannou část hry. Na základě toho se předpokládá, že šance hráčů získat bod na jejich podání v tomto modelu závisí pouze na jejich vlastní útočné síle a obranné síle jejich protihráčů. Uznává se, že tento předpoklad je idealizací, protože přijímající hráč může použít útočné prvky své hry při returnu nebo tam, kde se rozvine hra. Tento předpoklad se však používá ke snížení složitosti modelu. Pravděpodobnost, že hráč  $i$  získá bod proti hráči  $j$  při jejich podání, je definována takto:

$$P(r_{ij} = 1 | s_i, s_j) = \sigma(s_i - b_j)$$

Když  $r_{ij}$  nyní konkrétně odkazuje na pravděpodobnost, že hráč  $i$  získá bod při podání proti hráči  $j$ , místo toho, aby získal bod obecně. Všimněte si, že  $r_{ij} \neq 1 - r_{ij}$ , což by jinak platilo v obecném případě. Pravděpodobnost lze nyní vyjádřit jako:

$$P(D|S) = \prod_{k=1}^n [P(r_{ij}^k | s_i^k, b_j^k) P(r_{ij}^k | b_i^k, s_j^k)]$$

kde:

$$P(r_{ij}^k | s_i^k, b_j^k) = \sigma(s_i^k - b_j^k)^{r_{ij}^k} (1 - \sigma(s_i^k - b_j^k))^{(1-r_{ij}^k)}$$

Pokud máme danou množinu útočných a obranných dovedností dvou hráčů lze předpovědět probabilitu zisku bodu kterékoliv hráče při jeho podání. Markovův řetězec lze poté použít k převodu těchto pravděpodobností na pravděpodobnosti výhry.

### 1.13.10 Low-level point model

Tento model hledá cesty k zisku nebo ztrátě každého míčku v každém utkání sledovaných hráčů. Ke každé cestě přiřazuje určitou pravděpodobnost a pomocí Markovových řetězců určuje, jaká je pravděpodobnost zisku, či ztráty. Na následujícím obrázku můžeme vidět všechny události, které se při jedné výměně mezi tenisty mohou udát (Spanias & Knottenbelt, 2013).



P(vítězný bod) se pak tedy musí rovnat součtu všech podmíněných pravděpodobností, které vedou do bodu Vítězný bod =  $P(1SA) + P(1SR)P(1SRW|1SR) + P(1SF)P(2SA|1SF) + P(1SF)P(2SR|1SF)P(2SRW|2SR|1SF)$

P(ztracený bod) =  $P(1SR)P(1SRL|1SR) + P(1SF)P(2SR|1SF)P(2SRL|2SR|1SF) + P(1SF)P(2SF|1SF)$

V dalším kroku se spočtou pomocné proměnné reprezentující všechny uvažované zápasy jednoho hráče:

$p_{av}$  = průměrný počet míčků ve všech zápasech (dopsat pra)

$a = \frac{z}{P_{Av}}$        $z$  = počet es v zápase

$b$  = úspěšnost prvního podání v %

$c$  = zisk bodu z prvního podání v %

$d = \frac{y}{P_{Av}}$        $y$  = počet dvojchyb v zápase

$e$  = zisk bodu při druhém podání

$f = \frac{x}{P_{Av}}$        $x$  = počet es soupeře v zápase

$g$  = úspěšnost prvního podání soupeře v %

$h$  = zisk bodu při prvním podání soupeře v %

$i$  = zisk bodu při druhém podání soupeře v %

$j = \frac{w}{P_{Av}}$        $w$  = počet dvojchyb soupeře v zápase

Dále je třeba vypočítat průměrné pomocné proměnné  $a_{av}$  až  $e_{av}$  přes všechny hráče.

Následně je možné spočítat kombinované pravděpodobnosti, které budou brát v potaz statistiky hráčů  $A$  a  $B$  v předešlých zápasech turnaje. Spočítají se kombinované pomocné proměnné:

$$a_c = a_a + f_b - a_{av}$$

$$b_c = b_a + g_b - b_{av}$$

$$c_c = 1 + c_a - h_b - c_{av}$$

$$d_c = d_a + j_b - d_{av}$$

$$e_c = 1 + e_a - i_b - e_{av}$$

A z nich se dopočítají pravděpodobnosti kombinovaného modelu zápasu mezi hráči *A* a *B*:

$$P(1SA) = a_c$$

$$P(1SR) = b_c - a_c$$

$$P(1SRW \mid 1SR) = (b_c * c_c - a_c) / (b_c - a_c)$$

$$P(1SF) = 1 - b_c$$

$$P(2SA \mid 1SF) = 0$$

$$P(2SF \mid 1SF) = d_c$$

$$P(2SR \mid 1SF) = 1 - d_c$$

$$P(2SRW \mid 2SR \mid 1SF) = e_c$$

Pravděpodobnost vítězství hráče *A* je:

$$P(1SA) + P(1SR) * P(1SRW \mid 1SR) + P(1SF) * P(2SR \mid 1SF) * P(2SRW \mid 2SR \mid 1SF).$$



# CÍLE A ÚKOLY PRÁCE

## 1.14 Cíl výzkumného projektu

Cílem práce je vybrat existující model určeného pro predikci výsledku tenisového utkání a následně ho otestovat na vybraném tenisovém turnaji, v tomto případě Australian Open 2021. Dnešní modely pracují s velkým množstvím dat, které vypočítávají ve složitých vzorcích a málokdy jsou tato data dohledatelná a vzorce výpočtů pochopitelné pro širokou veřejnost. Mým cílem je upravit existující model tak, aby byl jednodušší a pochopitelný pro většinu sázejících. Sázející by měli při svém sázení k ruce jednoduchý model, který bude pracovat na základních a jednoduše dohledatelných datech. Tato data by se počítala v připravených a přehledných tabulkách. Sázející by tak mohli dostat nový pohled na zápas založený na datech, který následně využijí při rozhodnutí o tom, jakého hráče označí za vítěze.

Základním krokem bude vytvořit rešerši modelů pro predikci výsledků tenisových utkání. Z této rešerše následně vybrat jeden model, který bude založen na dohledatelných datech pro běžného sázejícího. Pokud se žádný nenajde, upravím jeden z modelů. V dnešní době se dá získat daleko více statistik, které jsou k použití ihned po zápase. Některé modely však byly vytvořeny v době, kdy tyto statistiky nebyly dohledatelné, případně ani neexistovaly. Dnes, kdy je vše elektronické se dají dohledat i statistiky, které před nedávnou dobou ani nebyly zveřejňovány.

Aplikovaný model budu porovnávat s použitým modelem vybrané sázkové kanceláře. Model sázkové kanceláře zde popsán nebude, protože si tyto kanceláře své predikční modely hlídají. Budeme se ale orientovat podle výše kurzů. Pokud bude na hráče například vypsán kurz 1, znamená to, že model sázkové kanceláře predikuje tohoto hráče jako jasného favorita. Podle těchto čísel budeme porovnávat, zda sázková kancelář odhadla vítěze utkání správně a s jakým kurzem. V ideálním případě nastane situace, v níž sázková kancelář předpoví vítěze utkání špatně a na hráče bude vypsán vysoký kurz. Zároveň náš model předpoví vítěze správně a náš hypotetický vklad se vrátí s větším ziskem. Dalším cílem této diplomové práce je prokázat finanční rentabilitu sázení a to vedením fiktivního účtu, na kterém při stejné výši vkladu na každé jedno utkání bude více peněz než na začátku. Nejdříve se bude sázet podle pravděpodobnosti, kterou nám určí model. To znamená, že vsadíme všechny hráče, které model predikuje jako vítěze. Dále budeme sázet podle výhodnosti, která je popsána v kapitole Výhodnost sázky.

Model budu porovnávat také se statistikami zápasu po odehrání. Zjistím tak odchylky predikčního modelu od reality, rovněž zjistím, jak moc a především ve kterých aspektech hry se model odchyluje.

## METODIKA PRÁCE

Abych mohl být zvolen některý z modelů, bylo nutné zjistit, zda lze sehnat veškeré statistiky hráčů, které jsou potřeba k výpočtům. V dnešní době ale není problém sehnat velký objem dat ke zpracování. Spíše jde o historická data a náročnost jejich dohledání. Například získání dlouhodobé úspěšnosti hráče na konkrétním turnaji a povrchu, případně proti určitému soupeři.

Vybraný model také musí být celý veřejný, aby byly jednoduše vysvětlené všechny kroky a postupy zpracování dat a výpočtů. Někteří si svůj postup hlídají a zveřejní například jen část modelu, ale nedá se zpracovat celý. Pro mou predikci výsledku tenisového utkání byl zvolen model Low-level point model. Tento model působí jednoduše na vkládání dat a počítání výsledků a dalších pomocných hodnot. Bohužel tento model počítá s tolika historickými daty, které jsou pro laika, kterému je tato práce určena, velmi složitě dohledatelná. Proto došlo k úpravě a v modelu sice zůstaly stejné vzorce výpočtů, ale používají se data od prvního kola do čtvrtfinále. Čtvrtfinálový zápas je první kolo, kde byl predikován vítěz utkání na základě dat získaných v kolech předešlých.

### 1.15 Základní soubor

První stovka hráčů nastupuje do hlavní části turnaje automaticky, dalších 128 hráčů žebříčku se utká v kvalifikaci o 16 míst. Pořadatelé turnaje mají také k dispozici divoké karty, které mohou udělit. Pokud se z turnaje odhlásí hráč, který je v první stovce, šanci dostává takzvaný šťastný poražený. Tedy hráč, který vypadl v kvalifikaci, ale díky uvolněnému místu v hlavním turnaji má možnost nastoupit. Tento hráč je vybrán z hráčů z kvalifikace buďto losem, či postavením v tenisovém žebříčku. Žebříček profesionálních tenistů zveřejňuje pravidelně Asociace tenisových profesionálů (ATP Tour, 2021).

V této práci byly základním souborem právě data získaná v průběhu prvních 4 kol turnaje, jelikož bylo čerpáno z dat získaných ke konkrétnímu hráči na turnaji a další vlastnosti hráče, jako je jeho jméno, národnost či umístění v žebříčku, nebyly relevantní.

### 1.16 Výběrový soubor

Hráč, který byl v pavouku ve čtvrtfinále umístěn nejvýše, byl označen číslem 1 a další hráči pak byli očíslováni postupně až do čísla 8. Výběrovým souborem k čtvrtfinálovému kolu byla všechna data těchto osmi hráčů ve všech předešlých kolech. V semifinále byla

aplikována stejná data, a navíc data ze čtvrtfinále. K finálovému zápasu byly kromě zmíněných dat využita data ze semifinále.

### **1.17 Využití modelu Low-level point**

Při tvorbě řešerše metod byla hledána metoda, která bude pochopitelná a jednoduše vysvětlitelná z matematického hlediska, ale také s ohledem na možnosti sběru dat. Pokud by byl zvolen model, který by byl sice jednoduchý, ale využíval by data, která nejsou nikde zveřejněna, nebylo by možné, byť jednoduchý model, vypočítat. Z toho důvodu byl zvolen Low-level point model, který byl následně upraven. Do tohoto modelu byly vkládány data pouze z jednoho turnaje. První 4 kola se nesázelo na utkání a sbírala se data do tabulek pro následující výpočty. První zápasy, které byly predikovány, byly od čtvrtfinále, aby se poznalo, jakou mají hráči formu, jak jim sedí povrch, na kterým se turnaj hraje, atp.

### **1.18 Sběr dat**

Z oficiální statistik tenisového turnaje Australian Open (Tennis Australia, 2021b) byly po každém zápase uloženy hodnoty výsledků do vytvořené tabulky. Všechny tyto tabulky jsou v přílohách číslo 1 až 8. K výpočtům byly potřeba tyto hodnoty:

Na podání

- Počet es
- Počet dvojchyb
- Úspěšnost 1. podání včetně esa
- Body po 1. podání včetně esa
- Body po 2. podání

Na příjmu

- Body po 1. podání soupeře
- Body po 2. podání soupeře

Body

- Vyhrané míče celkem

Z těchto dat jsem provedl výpočet pomocných proměnných pro každého hráče. Je důležité zmínit, že model počítá s nulovou pravděpodobností esa ze druhého podání.

I přesto, že v některých zápasech eso ze druhého podání padlo, bylo to většinou ve fázi, kdy bylo o vítězi rozhodnuto.

- $P_{av}$      $\emptyset$  celkového počtu míčků v předchozích zápasech

Autoři modelu počítali  $P_{av}$  jako průměrný počet bodů ve všech utkáních všech hráčů v celém turnaji. Já jsem pro potřeby této práce počítal pouze průměrný počet bodů daného hráče v daném turnaji od prvního kola.

- $z$      $\emptyset$  es v předchozích zápasech
- $y$      $\emptyset$  dvojchyb v předchozích zápasech
- $x$      $\emptyset$  es soupeře v předchozích zápasech
- $w$      $\emptyset$  dvojchyb soupeře v předchozích zápasech
- $a$      $\frac{z}{P_{Av}}$
- $b$      $\emptyset$  úspěšnost prvního podání
- $c$      $\emptyset$  bodů  $z$  prvního podání
- $d$      $\frac{y}{P_{Av}}$
- $e$      $\emptyset$  bodů ze druhého podání
- $f$      $\frac{x}{P_{Av}}$
- $g$      $\emptyset$  úspěšnosti prvního podání soupeře
- $h$      $\emptyset$  míčků získaných po prvním podání soupeře
- $i$      $\emptyset$  míčků získaných po druhém podání soupeře
- $j$      $\frac{w}{P_{Av}}$

Ke spočítání  $P_{av}$  se použil průměrný počet míčků v zápasech sledovaných hráčů od prvního zápasu turnaje,  $b$  odpovídá procentu úspěšnosti prvního podání,  $c$  procento výhry při prvním podání,  $d$  vyjadřuje počet vyhraných míčků z druhého podání ku  $P_{av}$ , stejný přístup byl použit pro proměnné  $g$ ,  $h$ ,  $i$ .

### 1.19 Pavouk turnaje

V následující tabulce je tzv. pavouk sledovaných hráčů, ve kterém jsou prezentovány pomocné výpočty (průměrné  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $e$ ), které budou použity v kapitole Statistiky zápasu. Jsou zde znázorněna všechna utkání hráčů, postoupivších do čtvrtfinále, od prvního do čtvrtého kola. Samozřejmě pak i semifinále a finále.

Hlavním účelem této tabulky je vytvoření si přehledu zápasů, výsledků, kurzů a porovnat je na jednom místě. Tabulka funguje i jako rozcestník, protože jména hráčů jsou vytvořena jako odkazy, přes které se dostaneme na statistiky hráče a jednoduše se tak ukáží čísla, která jsou potřeba. Stejně tak nás přesměrují k predikcím utkání.



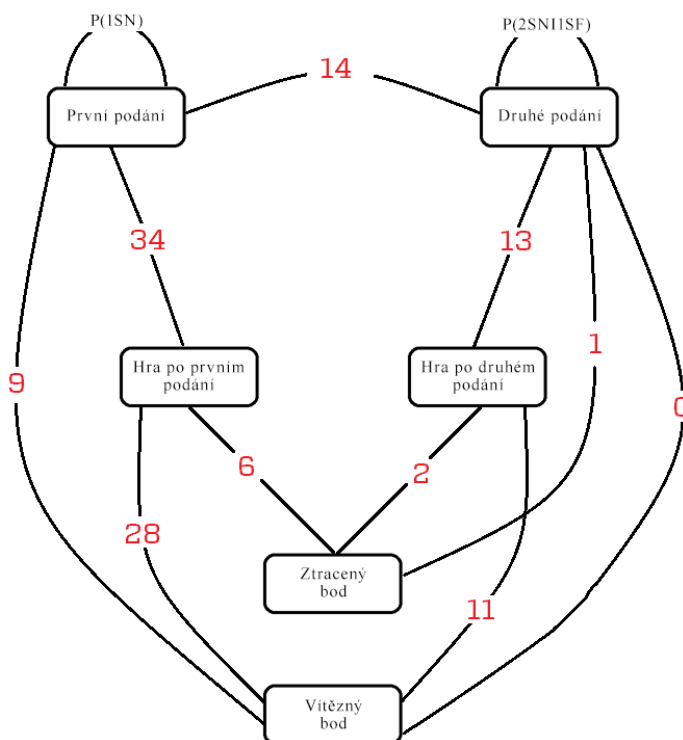
## VÝSLEDKY

Aby bylo možné udělat představu vyplněného Markovova řetězce, se v následujících kapitolách vybrali dva náhodné zápasy tohoto turnaje. Prvním z nich je zápas Djokoviče, který svůj zápas s Chardym zvládl velmi rychle a druhým je naopak zápas, který se hrál 5 setů.

Také jsou zde výsledky všech utkání od čtvrtfinále turnaje Australian Open 2021 a jejich porovnání s reálnými hodnotami zápasu získanými po odehrání turnaje.

### 1.20 Djokovič v zápase s Chardym

Jako ukázkou jsem si vybral dva zápasy. Níže je znázorněn průběh prvního kola zápasu, ve kterém si Djokovič poradil s Chardym lehce 3:0. Z grafu jsou patrné všechny statistiky podání, které pro naše výpočty potřebujeme. Tento graf bere v potaz pouze podávajícího hráče.



Obrázek 5 Statistika utkání Djokovič vs. Chardy při podání Djokoviče

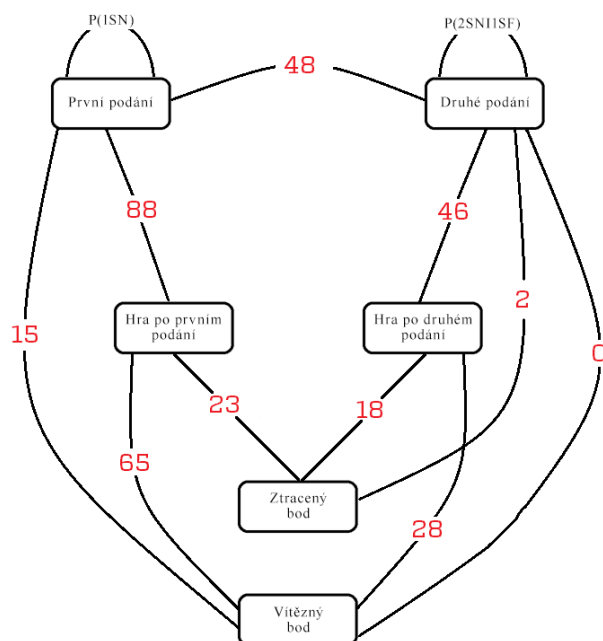
Je patrné, že Djokovič při svém podání soupeře nešetřil. Nadělil 9 es a pokud soupeř vybral první podání, tak byl Djokovič schopný uhrát 28 míčků a ztratil pouze 6. Čtrnáct prvních podání skončilo chybou, ale pouze jedenkrát skončilo dvojchybou. Druhé podání



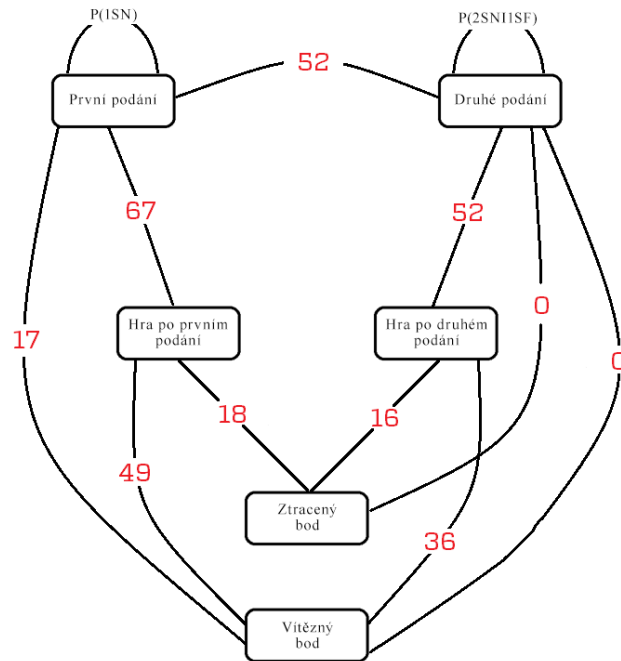
je většinou šance pro přijímacího hráče, protože je v naprosté většině pomalejší a hrané na „jistotu“. I tak Djokovič přetavil 11 druhých podání ve svůj prospěch a přišel pouze o dva míčky. Pokud spočítáme součty těchto řetězců, zjistíme, že Djokovič na svém podání uhrál  $9 + 28 + 11 + 0 = 48$  bodů a ztratil pouze  $6 + 2 + 1 = 9$  bodů. Když se podíváme zpátky do tabulky statistik, tak můžeme vidět, že tyto výpočty souhlasí.

## 1.21 Čtvrtfinálové utkání Nadal vs. Tsitsipas

Nevýhodou vybraných grafů je to, že neukazují vývoj zápasu. Jako příklad je možné uvést zápas mezi Nadalem a Tsitsipasem, ve kterém vedl Nadal už 2:0 na sety. Tsitsipas ale dokázal tento velmi kvalitní zápas otočit.



Obrázek 6 Statistika podání Nadala v zápase s Tsitsipasem



Obrázek 7 Statistika podání Tsitsipase v zápase s Nadalem

Pokud porovnáme grafy těchto dvou hráčů s Djokovičovým v předchozí kapitole, můžeme vidět, že zápas byl daleko vyrovnanější a déle trávající díky daleko větším číslům.

Nadal vyhrál  $15 + 65 + 28 + 0 = 108$  bodů na svém servise, což je dokonce o 6 míčků více než získal Tsitsipas, a přesto nepostoupil.

Pokud první podání bylo úspěšné (padlo do kurtu a soupeř odehrál), pak Nadal uhrál 65 z 88 míčků = 73 % a Tsitsipas 49 z 67 míčků = 73 %, což je neuvěřitelná shoda.

Druhé podání vyznělo lépe pro řeckého Tsitsipase, který uhrál 36 z 52 míčků = 69 %, zatímco Nadal uhrál 28 z 46 míčků = 60 %.

## 1.22 Statistiky zápasu

Predikována byla utkání od čtvrtfinále, na základě kterých se statistiky hráčů položily proti sobě. Nejdříve bylo nutné vypočítat průměr pomocných proměnných  $a_{av}$  až  $e_{av}$  ze statistik všech hráčů. Z těchto výpočtu byla získána kombinovaná statistika  $a_c$  až  $e_c$ , kterou poté bylo možné dosadit do vzorců pro výpočet kombinované pravděpodobnosti.

Tímto krokem byly dopočítány výsledné pravděpodobnosti výhry hráče  $A$ . Pro zjištění pravděpodobnosti výhry hráče  $B$  pak stačilo toto číslo odečíst od čísla 1.

V následujících tabulkách je také doplněn kurz na oba hráče, který byl platný těsně před začátkem utkání. Není to tedy základní kurz daný sázkařskou kanceláří, ale kurz, který se měnil od zveřejnění kurzu do začátku zápasu.

Pro ověření modelu v tabulce vznikly sloupce s názvem „Realita“ a „Odchylka“. Realita byla vypočítána po zápase, tedy v době, kdy byl znám vítěz a statistiky zápasu se staly dostupné. Odchylka je udána v absolutní hodnotě a ukazuje, jak se liší odhad modelu před zápasem k realitě po zápase.

Novak Djoković		VS.	Alexandr Zverev		Kolo 5					
ac	=	aA	+	fB	-	aav	=	0,02740		
		0,06205		0,02605		0,06070				
bc	=	bA	+	gB	-	bav	=	0,63406		
		0,67750		0,60500		0,64844				
cc	=	1	+	cA	-	hB	-	cav	=	0,65781
		1,00000		0,80750		0,34000		0,80969		
dc	=	dA	+	jB	-	dav	=	0,02640		
		0,01448		0,02730		0,01538				
ec	=	1	+	eA	-	iB	-	eav	=	0,65281
		1,00000		0,61750		0,37500		0,58969		
Pravděpodobnost		Vzorec		Low-level point model		Realita		Odchylka		
P(1SA)	=	ac	=	2,74%		14,20%		11,46%		
P(1SR)	=	bc - ac	=	60,67%		73,00%		12,33%		
P(1SRW 1SR)	=	(bc*cc-ac)/(bc-ac)	=	64,24%		89,58%		25,35%		
P(1SF)	=	1-bc	=	36,59%		27,00%		9,59%		
P(2SA 1SF)	=	0	=	0,00%		0,00%		0,00%		
P(2SF 1SF)	=	dc	=	2,64%		11,63%		8,99%		
P(2SR 1SF)	=	1-dc	=	97,36%		88,37%		8,99%		
P(2SRW 2SR 1SF)	=	ec	=	65,28%		50,00%		15,28%		
P(win)	=	P(1SA) + P(1SR) * P(1SRW 1SR) + P(1SF) * P(2SR 1SF) * P(2SRW 2SR 1SF)								
		0,649677293								
Hráč	Pravděpodobnost výhry	Výhodnost	Kurz	Výsledek	Vyšlo?					
Djoković	64,97%	4%	1,60	3	Ano					
Zverev	35,03%	-12%	2,50	1						

Tabulka 5 Pravděpodobnosti v zápase Djoković vs. Zverev

Čtvrtfinálový zápas Djokoviče a Zvereva, kteří patří mezi hráče z nejlepší desítky, někteří považovali za předčasné finále.

Bookmakeři favorizovali více srbského hráče, který vedl žebříček ATP.

Low-level point model také celkem jasně určil jako vítěze Djokoviče.

I přes výborný výkon německého hráče tento zápas dopadl podle očekávání. Zverev sice vyhrál první set v tie-braku, ale ve druhém a třetím setu nedostal od Djokoviče šanci.

Ve čtvrtém setu se hráči opět dostali do tie-breaku, který skončil ve prospěch Novaka a Zverev vypadl i přes vysoký počet es, který se zastavil na čísle 23.

Dimitrov Grigor		VS.	Karatsev Aslan		Kolo 5					
ac	=	aA	+	fb	-	aav	=	0,04282		
		0,07204		0,03147		0,06070				
bc	=	bA	+	gb	-	bav	=	0,66906		
		0,65250		0,66500		0,64844				
cc	=	1	+	cA	-	hb	-	cav	=	0,58781
		1		0,81250		0,41500		0,80969		
dc	=	dA	+	jb	-	dav	=	0,01237		
		0,01201		0,01574		0,01538				
ec	=	1	+	eA	-	ib	-	eav	=	0,49781
		1		0,56250		0,47500		0,58969		
Pravděpodobnost		Vzorec		Low-level point model		Realita		Odchylka		
P(1SA)	=	ac	=	4,28%		9,18%		4,90%		
P(1SR)	=	bc - ac	=	62,62%		64,00%		1,38%		
P(1SRW 1SR)	=	(bc*cc-ac)/(bc-ac)	=	55,96%		53,70%		2,26%		
P(1SF)	=	1-bc	=	33,09%		36,00%		2,91%		
P(2SA 1SF)	=	0	=	0,00%		0,00%		0,00%		
P(2SF 1SF)	=	dc	=	1,24%		20,00%		18,76%		
P(2SR 1SF)	=	1-dc	=	98,76%		80,00%		18,76%		
P(2SRW 2SR 1SF)	=	ec	=	49,78%		53,57%		3,79%		
P(win)	=	P(1SA) + P(1SR) * P(1SRW 1SR) + P(1SF) * P(2SR 1SF) * P(2SRW 2SR 1SF)								
		0,555990922								
Hráč	Pravděpodobnost výhry	Výhodnost	Kurz	Výsledek	Vyšlo?					
Dimitrov	55,60%	-26%	1,33	1						
Karatsev	44,40%	60%	3,6	3		Ne				

Tabulka 6 Pravděpodobnosti v zápase Dimitrov vs. Karatsev

Další čtvrtfinálový zápas spolu odehráli Karatsev a Dimitrov. Karatsev byl jediný účastník čtvrtfinálových bojů, který nebyl v nejlepší stovce hráčů ATP. I proto pro většinu fanoušků měl být tento zápas pro Dimitrova snadná cesta do semifinále. První set tomu byl důkazem, protože vyhrál 6:2.

Sázkové kanceláře měly Dimitrova za velkého favorita.

Low-level point model už nebyl tak výrazný, přeci jen 55 % není úplně jasná převaha.

Po prohraném prvním setu ale Karatsev předváděl výborný tenis, Dimitrova nepustil do hry a zbylé tři sety vyhrál. V tomto zápase se tedy nepovedlo predikovat vítěze ani v jednom případě.

Rublev Andrey		VS.	Medveděv Daniil		Kolo 5					
ac	=	aA	+	fb	-	aav	=	0,02585		
		0,07391		0,01264		0,06070				
bc	=	bA	+	gB	-	bav	=	0,55656		
		0,60750		0,59750		0,64844				
cc	=	1	+	cA	-	hB	-	Cav	=	0,64281
		1		0,83000		0,37750		0,80969		
dc	=	dA	+	jB	-	dav	=	0,00912		
		0,00905		0,01545		0,01538				
ec	=	1	+	eA	-	iB	-	eav	=	0,51031
		1		0,60250		0,50250		0,58969		
Pravděpodobnost		Vzorec		Low-level point model		Realita		Odchylka		
P(1SA)	=	ac	=	2,58%		9,41%		6,83%		
P(1SR)	=	bc - ac	=	53,07%		47,00%		6,07%		
P(1SRW 1SR)	=	(bc*cC-ac)/(bc-ac)	=	62,54%		53,13%		9,42%		
P(1SF)	=	1-bc	=	44,34%		53,00%		8,66%		
P(2SA 1SF)	=	0	=	0,00%		0,00%		0,00%		
P(2SF 1SF)	=	dc	=	0,91%		6,67%		5,75%		
P(2SR 1SF)	=	1-dc	=	99,09%		93,33%		5,75%		
P(2SRW 2SR 1SF)	=	ec	=	51,03%		59,52%		8,49%		
P(win)	=	P(1SA) + P(1SR) * P(1SRW 1SR) + P(1SF) * P(2SR 1SF) * P(2SRW 2SR 1SF)								
		0,581992936								
Hráč	Pravděpodobnost výhry	Výhodnost	Kurz	Výsledek	Vyšlo?					
Rublev	58,20%	78%	3,05	0	Ne					
Medvedev	41,80%	-41%	1,42	3						

Tabulka 7 Pravděpodobnosti v zápase Rublev vs. Medvedev

I v dalším zápase mezi Rublevem a Medveděvem se utkali dva hráči v top formě, a i přes jasný výsledek to byl, alespoň v prvních dvou setech, velmi vyrovnaný tenis.

Sázkové kanceláře více věřily mladému Medvedevovi.

Low-level point model ale za vítěze utkání určil Rubleva.

V tomto zápase však predikce modelu nevyšla a Medveděv tento zápas zvládl jednoduše 3:0 na sety.

Tsitsipas Stefanos		VS.	Nadal Rafael		Kolo 5					
ac	=	aA	+	fB	-	aav	=	0,02472		
		0,04200		0,04341		0,06070				
bc	=	bA	+	gB	-	bav	=	0,57406		
		0,61000		0,61250		0,64844				
cc	=	1	+	cA	-	hB	-	cav	=	0,64781
		1		0,82000		0,36250		0,80969		
dc	=	dA	+	jB	-	dav	=	0,03205		
		0,02262		0,02481		0,01538				
ec	=	1	+	eA	-	iB	-	eav	=	0,54281
		1		0,64000		0,50750		0,58969		
Pravděpodobnost		Vzorec		Low-level point model		Realita		Odchylka		
P(1SA)	=	ac	=	2,47%		12,50%		10,03%		
P(1SR)	=	bc - ac	=	54,93%		62,00%		7,07%		
P(1SRW 1SR)	=	$(bc \cdot cc - ac) / (bc - ac)$	=	63,20%		73,13%		9,94%		
P(1SF)	=	1 - bc	=	42,59%		38,00%		4,59%		
P(2SA 1SF)	=	0	=	0,00%		0,00%		0,00%		
P(2SF 1SF)	=	dc	=	3,20%		0,00%		3,20%		
P(2SR 1SF)	=	1 - dc	=	96,80%		100,00%		3,20%		
P(2SRW 2SR 1SF)	=	ec	=	54,28%		69,23%		14,95%		
P(win)	=	P(1SA) + P(1SR) * P(1SRW 1SR) + P(1SF) * P(2SR 1SF) * P(2SRW 2SR 1SF)							0,595680007	
Hráč		Pravděpodobnost výhry		Výhodnost		Kurz		Výsledek		Vyšlo?
Tsitsipas		59,57%		83%		3,07		3		Ano
Nadal		40,43%		-43%		1,42		2		

Tabulka 8 Pravděpodobnosti v zápase Tsitsipas vs. Nadal

V posledním zápase se potkal favorit na celkové vítězství Nadal s velkým talentem řeckého tenisu Tsitsipasem, který byl v roce 2021 ve vrcholné formě.

Málokdy se stane, že kurz na Nadala je větší než na soupeře, a ne jinak tomu bylo i zde a sázkové kanceláře favorizovali Nadala jako vítěze zápasu.

Zde ale Low-level point model docela překvapivě určil jako vítěze Tsitsipase.

Tento zápas trval přes čtyři hodiny a byla to bitva hodna finále turnaje. Nadal využil všech svých dovedností a soupeře v prvních dvou setech s přehledem přehrával. Kdekdo by si mohl myslet, že je výsledek zápasu již jistý. Avšak při tie-breaku třetího setu došlo ke zlomu a Nadal prohrál. Zápas tedy pokračoval. V dalším pokračování zápasu Nadalovi viditelně docházely síly, a i přes výbornou techniku mladšímu soupeři nestačil fyzicky a skončil před branou finále.

Djokovič Novak			VS.	Karatsev Aslan			Kolo 6			
ac	=	aa	+	fb	-	aav	=	0,03845		
		0,06614		0,03437		0,06205				
bc	=	ba	+	gb	-	bav	=	0,70075		
		0,68800		0,66000		0,64725				
cc	=	1	+	ca	-	ha	-	cav	=	0,57400
		1		0,79000		0,41200		0,80400		
dc	=	da	+	jb	-	dav	=	0,02000		
		0,01514		0,01996		0,01510				
ec	=	1	+	ea	-	ib	-	eav	=	0,54275
		1		0,58200		0,45400		0,58525		
Pravděpodobnost		Vzorec		Low-level point model		Realita		Odchylka		
P(1SA)	=	ac	=	3,85%		20,99%		17,14%		
P(1SR)	=	bc - ac	=	66,23%		68,00%		1,77%		
P(1SRW 1SR)	=	(bc*cc-ac)/(bc-ac)	=	54,93%		57,89%		2,97%		
P(1SF)	=	1-bc	=	29,93%		32,00%		2,08%		
P(2SA 1SF)	=	0	=	0,00%		0,00%		0,00%		
P(2SF 1SF)	=	dc	=	2,00%		7,69%		5,69%		
P(2SR 1SF)	=	1-dc	=	98,00%		92,31%		5,69%		
P(2SRW 2SR 1SF)	=	ec	=	54,28%		66,67%		12,39%		
P(win)	=	P(1SA) + P(1SR) * P(1SRW 1SR) + P(1SF) * P(2SR 1SF) * P(2SRW 2SR 1SF)								
		0,561400254								
Hráč	Pravděpodobnost výhry	Výhodnost	Kurz	Výsledek	Vyšlo?					
Djokovič	56,14%	-39%	1,08	3	Ano					
Karatsev	43,86%	100%	9,89	0						

Tabulka 9 Pravděpodobnosti v zápase Djokovič vs. Karatsev

Překvapení turnaje Karatsev narazil v semifinále na světovou jedničku Djokoviče.

Sázková kancelář určila jako jasného vítěze srbského hráče.

I low-level point model se přikláněl na stranu tohoto hráče s pravděpodobností 56,14 %.

Zápas měl jasný průběh a Djokovič vyhrál ve třech setech. I přes tato čísla nebyl kurz 1,08 na vítěze výhodný, což ukazuje další statistika. Podle čísel modelu byla sázka na Djokoviče velmi riskantní, za velmi nízký kurz. Za výhru Djokoviče pro sázce 100,- si sázející vydělal 8,-. Pokud by ale zvolil druhého hráče, mohl by vyhrát skoro devítinásobek vkladu.



Medveděv Daniil		VS.	Tsitsipas Stefanos		Kolo 6					
ac	=	aA	+	fb	-	aav	=	0,04415		
		0,06426		0,04194		0,06205				
bc	=	bA	+	gB	-	bav	=	0,65475		
		0,62200		0,68000		0,64725				
cc	=	1	+	cA	-	hB	-	Cav	=	0,64650
		1		0,79800		0,34750		0,80400		
dc	=	dA	+	jB	-	dav	=	0,02389		
		0,01691		0,02208		0,01510				
ec	=	1	+	eA	-	iB	-	eav	=	0,50225
		1		0,56000		0,47250		0,58525		
Pravděpodobnost		Vzorec		Low-level point model		Realita		Odchylka		
P(1SA)	=	ac	=	4,42%		20,73%		16,32%		
P(1SR)	=	bc - ac	=	61,06%		68,00%		6,94%		
P(1SRW 1SR)	=	$(bc * cc - ac) / (bc - ac)$	=	62,09%		82,05%		19,96%		
P(1SF)	=	1 - bc	=	34,53%		32,00%		2,53%		
P(2SA 1SF)	=	0	=	0,00%		0,00%		0,00%		
P(2SF 1SF)	=	dc	=	2,39%		7,69%		5,30%		
P(2SR 1SF)	=	1 - dc	=	97,61%		92,31%		5,30%		
P(2SRW 2SR 1SF)	=	ec	=	50,23%		50,00%		0,23%		
P(win)	=	P(1SA) + P(1SR) * P(1SRW 1SR) + P(1SF) * P(2SR 1SF) * P(2SRW 2SR 1SF)								
		0,592555151								
Hráč	Pravděpodobnost výhry	Výhodnost	Kurz	Výsledek	Vyšlo?					
Medvedev	59,26%	-16%	1,42	3	Ano					
Tsitsipas	40,74%	24%	3,04	0						

Tabulka 10 Pravděpodobnosti v zápase Medveděv vs. Tsitsipas

Ve druhém semifinálovém zápase se utkal ruský hráč Medveděv s Tsitsipasem, který porazil Nadala v předchozím kole.

Bookmakeři více věřili ruskému hráči v kurzu 1,42.

Low-level point model také předpověděl jako vítěze hráče z Ruska.

Na Tsitsipasovi se podepsalo utkání s Nadalem, které ho během pěti dlouhých setů stálo obrovské množství sil. Medveděvovi také pomohlo 17 es oproti soupeřovým 3. V této situaci předpověď vyšla v obou případech.



Djoković Novak		VS.	Medvedev Daniil		Kolo 7			
ac	=	aA	+	fb	-	aav	=	0,01679
		0,07133		0,01869		0,07323		
bc	=	bA	+	gB	-	bav	=	0,57231
		0,68667		0,59167		0,70603		
cc	=	1	+	CA	-	hB	-	Cav
		1		0,77667		0,38000		0,73050
								0,66617
dc	=	dA	+	jB	-	dav	=	0,01608
		0,01498		0,01589		0,01479		
ec	=	1	+	eA	-	iB	-	eav
		1		0,58833		0,47833		0,59000
								0,52000
Pravděpodobnost		Vzorec		Low-level point model		Realita		Odchylka
P(1SA)	=	ac	=	1,68%		3,85%		2,17%
P(1SR)	=	bc - ac	=	55,55%		67,00%		11,45%
P(1SRW 1SR)	=	$(bc * cc - ac) / (bc - ac)$	=	65,61%		71,43%		5,82%
P(1SF)	=	1 - bc	=	42,77%		33,00%		9,77%
P(2SA 1SF)	=	0	=	0,00%		0,00%		0,00%
P(2SF 1SF)	=	dc	=	1,61%		7,69%		6,08%
P(2SR 1SF)	=	1 - dc	=	98,39%		92,31%		6,08%
P(2SRW 2SR 1SF)	=	ec	=	52,00%		62,50%		10,50%
P(win)	=	$P(1SA) + P(1SR) * P(1SRW 1SR) + P(1SF) * P(2SR 1SF) * P(2SRW 2SR 1SF)$						
		0,600075845						
Hráč	Pravděpodobnost výhry	Výhodnost	Kurz	Výsledek	Vyšlo?			
Djoković	60,01%	12%	1,86	3	Ano			
Medvedev	39,99%	-18%	2,05	0				

Tabulka 11 Pravděpodobnosti v zápase Djoković vs. Medveděv

Do finále dokráčela světová jednička Novak Djoković relativně bez problémů. Ani Medveděv neměl moc složitou cestu. Oba hráči shodně ve třetím kole odehráli pětisetové bitvy. V ostatních zápasech dokonce Medveděv nedovolil soupeřům uhrát ani set. Finále tak slibovalo opravdové vyvrcholení celého turnaje a skvělý tenis.

Podle kurzů to měla být velmi vyrovnaná bitva.

Low-level point model favorizoval Djokoviče s pravděpodobností 60 %, což na finálový zápas není málo.

Ač tomu některé statistiky nenasvědčovaly (například to, že Medveděv vyhrál 44 % všech míčků), tak zápas skončil jasně 3:0 na sety pro Djokoviče, který tak oslavil další grandslamový titul.

## 1.23 Celkový přehled všech analyzovaných výsledků Australian Open 2021

Pokud se podíváme na celkové výsledky modelu uvedené v předchozí kapitole, pak je zřejmé, že model ze 7 utkání předpověděl u 5 utkání správného vítěze. Ve zbývajících dvou případech bychom ztratili svůj vklad. Pravděpodobnost určení správného hráče jako vítěze je v tomto případě tedy 72 %. V momentě, kdy model určí správného vítěze a sázkové kanceláře se netrefí, plyne nám z výhry velký zisk jako například v zápase Tsitsipas vs. Nadal, ve kterém byl kurz na vítěze utkání 3,07. V případě, že model určí vítěze správně a sázkové kanceláře stejného hráče mají za favorita, náš zisk bude zanedbatelný. Pokud model předpoví vítěze špatně, přicházíme o celý vklad.

Čtvrtfinále:

Djokovič vs. Zverev	zisk	0,60
Dimitrov vs. Karatsev	ztráta	1,00
Rublev vs. Medvedev	ztráta	1,00
Tsitsipas vs. Nadal	zisk	2,07

Semifinále:

Djokovič vs. Karatsev	zisk	0,08
Medvedev vs. Tsitsipas	zisk	0,42

Finále:

Djokovič vs. Medvedev	zisk	0,86
-----------------------	------	------

Pokud bychom na každý zápas vložili 100 korun českých, Celková vsazená částka by byla 700,-. Ztratili bychom peníze ve dvou utkáních a v ostatních získali. Celkem Bychom mělo po skončení turnaje na účtu 903,-.

Jinou variantou než sázet podle pravděpodobnosti modelu, je sázení podle výhodnosti sázky. Jak je popsáno v teoretické části, výhodnost se vypočítá jednoduchým vzorcem – pravděpodobnost výhry vynásobená kurzem a ponížena o jedna. Pokud by se sázelo podle tohoto ukazatele měli bychom na konci turnaje na účtě dokonce 1013,-

AO 2021												
<b>Čtvrtfinale</b>	Djokovic	Zverev	3:1	Dimitrov	Karatsev	1:3	Rublev	Medvedev	0:3	Tsitsepas	Nadal	3:2
Kurz	1,60	2,50		1,33	3,6		3,05	1,42		3,07	1,42	
Pravděpodobnost	64,97%	35,03%		55,60%	44,40%		58,20%	41,80%		59,57%	40,43%	
Výhodnost	3,95%	-12,42%		-26,05%	59,84%		77,51%	-40,64%		82,87%	-42,59%	
<b>Semifinale</b>	Djokovic	Karatsev	3:0	Medvedev	Tsitsepas	3:0						
Kurz	1,08	9,89		1,42	3,04							
Pravděpodobnost	56,14%	43,86%		59,26%	40,74%							
Výhodnost	-39,37%	100,00%		-15,86%	23,86%							
<b>Finále</b>	Djokovic	Medvedev	3:0									
Kurz	1,86	2,05										
Pravděpodobnost	60,01%	39,99%										
Sázka	11,61%	-18,02%										
			Počáteční kapitál	700								
			Konečný kapitál									
			Podle pravděpodobnost	903								
			Podle výhodnosti kurzu	1013								

Tabulka 12 Celkový přehled výsledků Australian Open 2021 s pravděpodobnostmi, kurzy, sázky a výhodností kurzu

Z čísel je vidět, že bychom byli úspěšnější a na konci turnaje by na účtu zůstalo více peněz. Při sázení podle výhodnosti totiž eliminujeme sázení na velmi nízké kurzy, které jsou nevýhodné, protože v případě výhry získáme málo a v případě prohry přijdeme stále o stejnou výši vkladu. Konkrétně na těchto výsledcích jsme nevsadili na Djokoviče v semifinálovém utkání s Karatsevem. Sice jsme přišli o vklad, ale kdyby tento zápas vyšel, zisk by byl velký. To se stalo v případě zápasu Dimitrova s Karatsevem, kde jsme nevsadili na pravděpodobnějšího vítěze podle predikce modelu, ale na méně pravděpodobnějšího vítěze. Získali jsme tak ve svůj prospěch kurz 3,6.

## 1.24 Výsledek predikce dalších vybraných grandslamů

Výsledky predikcí na turnaji v Australian Open mě přiměly k vypracování tabulek pro další grandslamy v roce 2022. Hlavně z důvodu eliminace faktoru náhody. Předěšlý turnaj vyšel v zelených číslech. Mohlo to být ale štěstí a dalších několik turnajů by vyšlo záporně. To by z dlouhodobého hlediska znamenalo velké ztráty při použití modelu, přestože byl na Australian Open úspěšný.

Při zpracovávání dat došlo k poučení se z některých zbytečných kroků a následující tabulky se díky tomu tvořily opravdu rychle, protože již bylo jasné, jaká data jsou potřeba a nebyla přepisována zbytečná data, které se do výpočtů nepoužívala.

Také výpočty excelových tabulek se posunuly. Například karty nebyly pojmenovány podle hráčů, ale podle svislého pořadí v pavoukovi. Hráč číslo 1 byl tedy v pavouku postaven nejvýše a hráč číslo 8 nejnižší. V této fázi se tedy vždy museli potkat hráči v daném pořadí. Další zápasy byly označeny jako 1v2, 3v4, 5v6 a 7v8. V dalším zápase se utkal někdo z hráčů 1 nebo 2 proti hráči 3 nebo 4. Je pro nás tedy irelevantní, jak se hráč jmenuje, stačí nám data, která v průběhu turnaje nasbírání. Pro mě to tedy znamenalo, že jsem počítal každého hráče jako by se dostal do finále. Hráč, který vypadl, měl na těchto místech prázdná pole, a proto výsledky výpočtů nijak neovlivňoval. Tabulka semifinálových utkání se tedy po zadání hráčských statistik z minulých kol dokázala spočítat automaticky. Následné porovnávací výpočty už se také dopočítávali po zadání základních údajů z tenisových klání dotyčných hráčů. Tento systém má obrovskou výhodu v rychlosti zadávání a počítání. Naopak nevýhodou je následné upravování modelu. Při přemýšlení, jak model vylepšit, s jakým dalším údajem počítat, jsem omezený na vypsání hodnoty, které jsou potřeba k tomuto modelu. Například pokud bych

chtěl do výpočtů zahrnout i počet vyhraných nebo odvrácených breakballů, musím data pro tento údaj do tabulek doplňovat a nemohu ho rovnou použít.

Následující obrázek ukazuje upravenou tabulku hráče, ve které jsou čísla nutná k následujícím výpočtům.

Jméno	Fokina					Výpočty
<b>Podání</b>	<b>Počet</b>	<b>Počet</b>	<b>Počet</b>	<b>Počet</b>		Pav 274,00
Esa	8	3	2	2		z 3,75
Dvojchyby	3	5	4	3		y 3,75
Úspěšnost 1. podání včetně esa	63%	65%	68%	65%		x 4,25
Body po 1. podání včetně esa	74%	67%	64%	70%		w 6,00
Body po 2. podání	57%	52%	38%	49%		a 0,01
<b>Příjem</b>						b 65%
Body po 1. podání soupeře	38%	34%	37%	35%		c 69%
Body po 2. podání soupeře	57%	50%	42%	54%		d 0,0137
<b>Body</b>						e 49%
Míče celkem	175	318	361	242		f 2%
<b>Soupeř</b>						g 61%
Esa	5	5	6	1		h 36,00%
Dvojchyby	4	9	6	5		i 50,75%
Úspěšnost 1. podání	58%	62%	64%	61%		j 2%

Tabulka 13 Statistiky prvních 4 kol hráče: Fokina

Na první pohled je vidět, že tabulka je daleko chudší na stažená data. Zadávaní dat do tabulky se tím zrychlilo na třetinu času. Tabulka pavouka zůstává stejná, protože na ní není z mého pohledu nic k odstraňování a slouží k celkovému přehledu soutěže a k mezi výpočtům. Ani na dalších listech výpočtů nedošlo k velkým změnám.

#### **1.24.1 Australian Open 2022**

Jak znázorňuje tabulka 14 na následující straně, tento turnaj nevyšel v zelených číslech. Můžeme být sice spokojeni s počtem odhadnutých vítězů. Nicméně sázka s tak nízkým kurzem nám nedokázala dorovnat ani prodělek ve výši dvou ztrátových zápasů, přestože bylo správně předpovězeno 5 ze 7 zápasů. To znamená, že i s úspěšností 71 % se nedokázal model dostat do zisku, ale naopak jsem skončil ve ztrátě. Pokud by se sázelo podle výhodnosti kurzu, bylo by množství peněz po skončení turnaje na účtě ještě menší.

Data pro porovnání byla získána z oficiálních webových stránek ATP Tour (ATP Tour, 2022).

AO2022																											
<b>Čtvrtfinale</b>	Djoković	Beretini	3	:	1	Nadal	Schwartzman	3	:	1	Zverev	Fokina	3	:	0	Tsitšipas	Medvedev	3	:	0							
Kurz	1,15	6,16				1,06	12				1,12	7,05				1,41	3,12										
Pravděpodobnost	60,34%	39,66%				47,38%	52,62%				57,47%	42,53%				52,42%	47,58%										
Výhodnost	-30,60%	144,28%				-49,78%	100,00%				-35,63%	100,00%				-26,08%	48,44%										
<b>Semifinale</b>	Djoković	Nadal	3	:	1	Zverev	Tsitšipas	2	:	3																	
Kurz	2,83	1,48				2,97	1,44																				
Pravděpodobnost	41,64%	58,36%				36,92%	63,08%																				
Výhodnost	17,83%	-13,62%				9,66%	-9,17%																				
<b>Finále</b>	Djoković	Tsitšipas	3	:	2																						
Kurz	1,44	2,98																									
Pravděpodobnost	59,63%	40,37%																									
Výhodnost	-14,13%	20,30%																									
												<b>Počáteční kapitál</b>		700													
												<b>Konečný kapitál</b>															
												<b>Podle pravděpodobnosti</b>		626													
												<b>Podle výhodnosti kurzů</b>		383													

Tabulka 14 Celkový přehled výsledků Australian Open 2022 s pravděpodobnostmi, kurzy a sázkou

### 1.24.2 United States Open 2022

Další turnaj, který jsem si vybral pro tuto metodu je americký grandslam. Data pro porovnání byla čerpána z oficiálních webových stránek US Open (US Open, 2022). Náš model na tomto turnaji předpověděl vítěze zápasu pouze dvakrát ze sedmi utkání. To se projevilo i na financích. Zisk z dvou utkání byl 106 korun českých, ale ztráta 500 korun českých. Za zmínku stojí například utkání mezi Berrettinim a Ruudem, který náš model viděl jasně ve prospěch Berrettinyho, ale on prohrál 0:3.

Pokud by se sázelo podle výhodnosti kurzu, konto by bylo také po skončení turnaje menší než na začátku. I přestože vyšla sázka na první utkání mezi Kirgiosem a Khachanovem v kurzu 4,21, poté už nevyšel žádný zápas.



US2022																
<b>Čtvrtfinále</b>	Kyrgos	Khachanov	2	: 3	Berrettini	Rud	0	: 3	Sinner	Alcaraz	2	: 3	Rublev	Tarfoe	0	: 3
Kurz	1,26	4,21			1,87	2,03			2,53	1,58			1,94	2,02		
Pravděpodobnost	61,32%	38,68%			66,06%	33,94%			46,33%	53,67%			57,18%	42,82%		
Výhoda	-22,74%	62,86%			23,54%	-31,11%			17,22%	-15,21%			10,94%	-13,51%		
<b>Semi-finále</b>	Khachanov	Rud	1	: 3	Alcaraz	Tarfoe	3	: 2								
Kurz	3,07	1,46			1,48	2,98										
Pravděpodobnost	62,60%	37,40%			55,01%	44,99%										
Výhoda	92,20%	-45,40%			-18,58%	34,07%										
<b>Finále</b>	Rud	Alcaraz	1	: 3												
Kurz	2,96	1,49														
Pravděpodobnost	50,94%	49,06%														
Výhoda	51%	-27%														
<b>Počáteční kapitál</b>												700				
<b>Konečný kapitál</b>																
<b>Podle pravděpodobnosti</b>												306				
<b>Podle výhody kurzů</b>												421				

Tabulka 15 Celkový přehled výsledků United States Open 2021 s pravděpodobnostmi, kurzy a sázky

## DISKUZE

Tyto velké turnaje jsem vybral záměrně, protože na těchto turnajích jsou vypsané obrovské odměny za postup do dalšího kola, jak je popsáno v kapitole 1.4. Předpokládám tedy, že hráči budou daleko více motivováni postoupit do dalšího kola než k prodeji svého zápasu, a tedy ke zmanipulování sázek. K těmto podvodům jsou náchylnější hráči nižších kategorií. Pokud je například ve finále kurz na outsidera 3 a já, jako papírově silnější hráč, dostanu za vítězství v turnaji sto tisíc, je pro mě lehčí se se soupeřem domluvit, nechat ho vyhrát a vsadit si na něj tak, aby výhra v sázce byla daleko vyšší než v turnaji. Samozřejmě jako hráč mohu kurz upravit a třeba první set vyhrát, abych zvedl kurz na outsidera třeba na 5. Zajisté zde také hraje počet bodů a možnost tenisty se dostat do lépe hodnocených turnajů. Budu tedy předpokládat, že turnaj Grand Slam je tak prestižní a finančně ohodnocený, že hráči nebudou své zápasy prodávat. Už jen za první kolo si tenisté na účet připsou přes milion korun českých korun.

### 1.25 Porovnání výsledků modelu a reality na Australian Open 2021

Komparace výsledků predikce a reálných výsledků po odehrání daného zápasu je zobrazena v následující tabulce číslo 16. Toto porovnání bylo zvoleno z důvodu porovnání výsledku modelu v jednotlivých pravděpodobnostech

Nejmenší odchylka vznikla u pravděpodobnosti esa ze druhého podání  $P(2SA|1SF)$ . Přesně 0 % je ale dáno tím, že tato pravděpodobnost byla stanovena již na začátku práce. Při detailnější zkoumání tenisových utkání se sice některým hráčům povedlo zahrát eso ze druhého podání, ale nikdy tato situace nenastala za vyrovnaného stavu.

U pravděpodobnosti chyby na prvním podání  $P(1SF)$  se model ani v jednom případě nedostal nad 10 % odchylku a v průměru se rovná 1,83 %. Do jisté míry se tedy dá říct, že první podání a jeho pravděpodobnost úspěchu či neúspěchu neovlivňuje soupeř, ale hráč samotný, protože první podání bude chtít vždy riskovat na maximum. Samotný počet samozřejmě závisí na délce zápasu, ale pravděpodobnosti se moc neliší.

Hra, která neskončí esem ani zkaženým podáním  $P(1SR)$  má celkovou odchylku 4,98 % a ve dvou zápasech dokonce nedosáhla ani na 2 %. Point model obecně tuto pravděpodobnost počítá nižší, než je skutečná realita, která se průměrně pohybuje kolem 64 %.

Některá podání hráčů přesahují rychlost i přes 200 km/h. Přináší jim to obrovskou výhodu, pokud umí takovéto podání zahrát stabilně. Odchylka modelu od reality je v případě P(1SA) ve dvou zápasech dokonce větší než 15 %. V zápase Djokoviče s Karatsevem model předpověděl 4% pravděpodobnost esa v zápase, ale realita byla 21 %. I přes jasný výsledek a odehraných 27 gamů hráči zahráli 23 es.

Největší nepřesnost modelu je u pravděpodobnosti P(1SRW|1SR), která říká, že podávající hráč uhrál bod z prvního podání bez započítání es. Tato odchylka jako jediná přesáhla průměrně 10 %. V zápase Djokoviče se Zverevem byla dokonce přes 25 %. Nutno podotknout, že tento zápas statisticky v této pravděpodobnosti vybočil, protože hráči takto uhráli 90 % míčků, což je reálně 15 % nad průměrem.

Pokud hráč pokazí první podání, má ještě druhý pokus. Odchylka u P(2SF|1SF) a P(2SR|1SF) se rovná, protože bylo stanoveno, že ze druhého podání nepadne eso. Jelikož jsou pouze dvě cesty od 1SF – zkažené podání nebo navazující hra (z Markového řetězce je to dobře patrné), tak součet P(2SF|1SF) a P(2SR|1SF) se rovná jedné. Odchylka v tomto případě byla průměrně kolem 7 %. Pokud se hráči dostali do hry po druhém podání, pak v průměru vyhrával podávající hráč v 59 %. Toto číslo také ukazuje výhodu podání. Pokud ho porovnáme, tak hráči na prvním podání (bez započítaných es) měli v průměru 75% úspěšnost zisku bodu oproti zmíněným 59 % ze druhého podání.

	Djokovic	vs.	Zverev	Dimitrov	vs.	Karatsev
Pravděpodobnost	Point model	Realita	Odchylka	Point model	Realita	Odchylka
P(1SA)	2,74%	14,20%	11,46%	4,28%	9,18%	4,90%
P(1SR)	60,67%	73,00%	12,33%	62,62%	64,00%	1,38%
P(1SRW 1SR)	64,24%	89,58%	25,35%	55,96%	70,37%	14,41%
P(1SF)	36,59%	27,00%	9,59%	33,09%	36,00%	2,91%
P(2SA 1SF)	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
P(2SF 1SF)	2,64%	11,63%	8,99%	1,24%	20,00%	18,76%
P(2SR 1SF)	97,36%	88,37%	8,99%	98,76%	80,00%	18,76%
P(2SRW 2SR 1SF)	65,28%	50,00%	15,28%	49,78%	53,57%	3,79%

	Rublev	vs.	Medvedev	Tsitsipas	vs.	Nadal
Pravděpodobnost	Point model	Realita	Odchylka	Point model	Realita	Odchylka
P(1SA)	2,58%	9,41%	6,83%	2,47%	12,50%	10,03%
P(1SR)	53,07%	47,00%	6,07%	54,93%	62,00%	7,07%
P(1SRW 1SR)	62,54%	78,13%	15,58%	63,20%	73,13%	9,94%
P(1SF)	44,34%	53,00%	8,66%	42,59%	38,00%	4,59%
P(2SA 1SF)	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
P(2SF 1SF)	0,91%	6,67%	5,75%	3,20%	0,00%	3,20%
P(2SR 1SF)	99,09%	93,33%	5,75%	96,80%	100,00%	3,20%
P(2SRW 2SR 1SF)	51,03%	59,52%	8,49%	54,28%	69,23%	14,95%

	Djokovic	vs.	Karatsev	Medvedev	vs.	Tsitsipas
Pravděpodobnost	Point model	Realita	Odchylka	Point model	Realita	Odchylka
P(1SA)	3,85%	20,99%	17,14%	4,42%	20,73%	16,32%
P(1SR)	66,23%	68,00%	1,77%	61,06%	68,00%	6,94%
P(1SRW 1SR)	54,93%	57,89%	2,97%	62,09%	82,05%	19,96%
P(1SF)	29,93%	32,00%	2,08%	34,53%	32,00%	2,53%
P(2SA 1SF)	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
P(2SF 1SF)	2,00%	7,69%	5,69%	2,39%	7,69%	5,30%
P(2SR 1SF)	98,00%	92,31%	5,69%	97,61%	92,31%	5,30%
P(2SRW 2SR 1SF)	54,28%	66,67%	12,39%	50,23%	50,00%	0,23%

	Djokovic	vs.	Medvedev	Průměrné hodnoty		
Pravděpodobnost	Point model	Realita	Odchylka	Point model	Realita	Odchylka
P(1SA)	1,68%	3,85%	2,17%	3,15%	12,98%	9,83%
P(1SR)	55,55%	67,00%	11,45%	59,16%	64,14%	4,98%
P(1SRW 1SR)	65,61%	71,43%	5,82%	61,22%	74,66%	13,43%
P(1SF)	42,77%	33,00%	9,77%	37,69%	35,86%	1,83%
P(2SA 1SF)	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
P(2SF 1SF)	1,61%	7,69%	6,08%	2,00%	8,77%	6,77%
P(2SR 1SF)	98,39%	92,31%	6,08%	98,00%	91,23%	6,77%
P(2SRW 2SR 1SF)	52,00%	62,50%	10,50%	53,84%	58,78%	4,95%

Tabulka 16 Tabulka porovnání pravděpodobností a odchylek modelu od reality

## 1.26 Porovnání pravděpodobností modelu a sázkových kanceláří

Přepočet kurzů je popsán v kapitole Převod kurzů na pravděpodobnosti. V následující tabulce je modře označen vítěz utkání. Zeleně pokud model nebo sázková kancelář predikovala vítěze správně a červeně, pokud výsledek utkání byl jiný než predikce.

AO2021			
Model versus sázkovky	Djokovic	vs.	Zverev
Low-level point model	65%		35%
Kurz	1,60		2,50
Pravděpodobnost kurzu	65%		35%
Výhodnost kurzu	4%		-12%
	Dimitrov	vs.	Karatsev
Low-level point model	56%		44%
Kurz	1,33		3,60
Pravděpodobnost kurzu	81%		19%
Výhodnost kurzu	-26%		60%
	Rublev	vs.	Medveděv
Low-level point model	58%		42%
Kurz	3,05		1,42
Pravděpodobnost kurzu	25%		75%
Výhodnost kurzu	78%		-41%
	Tsitsipas	vs.	Nadal
Low-level point model	60%		40%
Kurz	3,07		1,42
Pravděpodobnost kurzu	25%		75%
Výhodnost kurzu	83%		-43%
	Djokovic	vs.	Karatsev
Low-level point model	56%		44%
Kurz	1,08		9,89
Pravděpodobnost kurzu	95%		5%
Výhodnost kurzu	-39%		100%
	Medveděv	vs.	Tsitsipas
Low-level point model	59%		41%
Kurz	1,42		3,04
Pravděpodobnost kurzu	75%		25%
Výhodnost kurzu	-16%		24%
	Djokovic	vs.	Medveděv
Low-level point model	60%		40%
Kurz	1,86		2,05
Pravděpodobnost kurzu	51%		49%
Výhodnost kurzu	12%		-18%

Tabulka 17 Celkové porovnání pravděpodobností modelu a kurzu

Ze sedmi utkání se predikce shodují v pěti utkáních s tím, že ve čtyřech z nich byly úspěšné. Ve dvou případech se tedy modely neshodují a pokaždé se prokázal jiný.

V zápase mezi Djokovičem a Zverevem jsou pravděpodobnosti velmi vyrovnané – 65 %. V zápase Djokovič vs. Karatsev a Medveděv vs. Tsitsipas byla pravděpodobnost výhry menší u Low-point modelu než u sázkových kanceláří, ale naopak tomu bylo ve finálovém zápase Djokovič vs. Medveděv. V jediném zápase, který ani jedna pravděpodobnost nedokázala správně určit, Dimitrov vs. Karatsev, si byl kurz na 81 % jistý výhrou Dimitrova, který však neuspěl. Model tak jednoznačný nebyl a tomuto hráči věřil pouze na 56 %.

V několikrát zmiňovaném zápase Tsitsipas vs. Nadal byl model přesnější a určil správně vítěze utkání s 60% pravděpodobností, zatímco sázkové kanceláře věřily více Nadalovi a to dokonce s pravděpodobností 75 %. Naopak tomu bylo v zápase RUBLEVA s Medveděvem, ve kterém model určil nesprávně vítěze s pravděpodobností 42 % a sázková kancelář si byla jistá na 75 %.

## ZÁVĚR

Při předchozím sledování kurzů náhodných utkání v sázkové kanceláři a sledování statistik bylo zjištěno, že se sázková kancelář nedokáže trefit z deseti případů devětkrát na vítěze v kurzu 1,5 – 1,6. Při dalším sledování v jiném časovém období se ale toto číslo obrátilo a sázková kancelář určila vítěze před utkáním v kurzu 1,5 – 1,6 devětkrát v řadě. Tato statistika je jasný impuls k tomu, že sázková kancelář nemá vítěze „pod kontrolou“ a jejich predikční modely nejsou spolehlivé. Samozřejmě kdyby byly, tak by lidé sázeli na ty hráče, kteří mají menší kurz a pokaždé by vyhrávali a sázkové kanceláře by neměly žádný zisk. Naopak by byly ve ztrátě a tento obrovský byznys by nefungoval.

Obecně platí, že vrcholový sport je velmi vyrovnaný a před zápasem si nemůže být nikdo jistý tím, kdo vyhraje. Každý, kdo tento sport sleduje ví, že tenista je schopný odehrát jeden zápas jako nikdy a poté další zápas prohrát, byť podle čísel s horším soupeřem. Roli zde může hrát několik faktorů jako například podcenění soupeře, či únava, která se může na těchto turnajích projevit, protože zápas může trvat i několik hodin a hráč odehraje několik těchto těžkých zápasů v krátké době po sobě. Z tohoto důvodu je z dlouhodobého hlediska nevýhodné sázet na kurzy které nemají hodnotu. To znamená riskovat vsazenou částku kvůli desetině výhry. Pokud sázející bude vsázet tento kurz (např. 1,1), musí trefit deset zápasů, aby se dostal alespoň na dvojnásobek a následná prohra by ho nedostala do záporných čísel. Tato strategie může chvíli vycházet, ale pak mohou přijít například dvě prohry za sebou a hráč se bude těžko z této situace dostávat. Hodnotné kurzy jsou ty, které se z dlouhodobého hlediska vyplatí i při několika neúspěších. Pokud sázející bude vyhledávat outsidersy, kteří mají šanci utkání vyhrát a je na ně vyšší kurz, například 3,5, stačí mu predikovat správně jedno ze tří utkání a dostane se do zisku.

Turnaj Australian Open 2021 byl vybrán jako první „jistý“, který se odehraje po coronavirové pandemii, s ohledem na naše přípravy a nastudování potřebných materiálů. Byla to nepředvídatelná doba, a ne každý měl stejné podmínky pro trénink. Navíc tento pandemický stav každý hráč snášel jinak. I proto byly do práce zahrnuty výpočty z dalších turnajů, aby se mohla porovnat data z více turnajů a odstranit tak faktor náhody. Díky počítání predikce pouze z dat utkání daného turnaje bylo možné potřebná data vložit do tabulky během pár minut.

Do zvoleného modelu byly použity statistiky hráčů pouze z prvních kol turnaje, kteří se dostali do čtvrtfinále. To sloužilo k odhadnutí aktuální formy hráčů v turnaji

a eliminování co nejvíce vlivů na tyto hráče. Není to ale tak jednoduché, protože ve statistikách nejsou zapsané všechny situace, kterými si hráč ve své kariéře prošel. Například pokud hráč ve své kariéře nikdy nedokázal proměnit stav 5:0 do vítězného konce, statistiky, které se zajímají o dlouhodobé údaje, by s touto možností mohly počítat. Pokud tato situace nenastala v předešlých čtyřech utkáních na turnaji, je to nová situace a upravený model by si s ní nedokázal poradit. Ze zkušenosti ale také vím, že psychická pohoda hráče se může změnit v okamžiku a hráč sice mohl zahrát výborně předchozí zápasy, ale ten další již může být ztrátový v důsledku stresující informace. I přestože jsou tyto vrcholoví sportovci psychicky odolní a připravují se na zvládnání těchto situací, přeci jen mohou nastat skutečnosti, které dokáží ovlivnit každého. Kdo tento sport sleduje, tak sám ví, že stejný tenista dokáže zahrát utkání jako nikdy předtím a v dalším kole, kde se utká s jiným hráčem, o kterém si myslíte, že do dalšího kola postoupil šťastnou náhodou, úplně jasně prohraje.

Z výsledků vyplývá, že model úspěšný být může. Může predikovat vítěze utkání se štěstím i z 80 %. Problém nastává s výší kurzů sázkových kanceláří. V naprosté většině model predikoval stejného vítěze jako sázková kancelář. To znamená, že na vítěze byl menší kurz a zisk by byl tedy malý, méně jak 40 %. Naopak v případě ztráty byla ztráta 100 % vkladu. Řešením by mohlo být sázet pouze tehdy, pokud půjde model proti předpovědi sázkové kanceláře a výhra by tak byla vyšší. Kdyby nastala shoda upraveného modelu se sázkovou kanceláří a kurzem například 1,05 sázející by získal v případě úspěchu 5 % vkladu. Toto číslo je z dlouhodobého hlediska ale velmi nevýhodné, ač by se mohlo zdát, že je to velká jistota. Tato sázka by musela vyjít ve dvaceti případech, aby se vklad vrátil. Pokud by však alespoň jednou, i přes tuto vysokou pravděpodobnost, predikce nevyšla, dostane se do ztráty.

Z tohoto důvodu byla přidána sázka podle výhodnosti kurzu, která je popsána v kapitole 2.8 Výhodnost sázky. Pokud model predikuje hráče s kurzem 1,1 s pravděpodobností 52 %, je zcela jistě lepší zkusit vsadit na druhého hráče, sice s o pár procenty nižší pravděpodobností, za to s daleko vyšším možným ziskem, v tomto případě i devítinásobným.

První turnaj dostal model do zisku, hlavně protože model správně predikoval vítěze zápasu Tsitsipas vs. Nadal, který sázkové kanceláře viděly obráceně. Vklad sázejícího se tedy ztrojnásobil a dokázal pokrýt ostatní ztráty. Nebýt tohoto utkání, byl by model ve ztrátě. Pokud by se sázelo podle výhodnosti, toto utkání by bylo predikováno také



správně. To vede k myšlence, že za ziskem stála spíše náhoda. Příkladem toho jsou statistiky z dalšího turnaje Australian Open 2022, kde bylo správně predikováno pět utkání ze sedmi a stejně by sázející skončil ve ztrátě 44 %. Tento výsledek potvrzuje, že predikovat vítěze utkání není jednoduché, a i přes větší úspěšnost správného odhadnutí vítěze jsou sázky nastavené tak, že z dlouhodobého hlediska vždy vyhraje sázková kancelář. Pokud by se zde sázelo podle výhodnosti, na účtě by zbylo jen něco málo přes polovinu našeho vkladu do hry.

Existují hráči, kterým se povedlo vyhrát větší peníze, ale do větší míry je to dílo náhody a velkého štěstí. Na takové hráče připadá velké množství sázejících, kteří jsou ve ztrátě, a proto může sázkařský byznys růst. Myslím si, že finanční zatížení státem by mohlo být daleko větší a peníze by měly jít cíleně na rozvoj sportovní infrastruktury a mládeže a pomoci tak ke sportu přilákat větší množství dětí. To vše podpořit materiálním vybavením a vzděláním trenérů.

Závěrem je tedy potřeba říci, že sázení je velmi riskantní záležitost, která by nikdy neměla přerůst v gamblersství. Už jen protože na pravděpodobnost 50 % je vypsán kurz 1,85 je velmi těžké sázkovou kancelář porazit. Pokud hráč vsadí na zápas například 5 korun jen proto, aby pro něj byl zápas emotivnější a měl z něj větší prožitek, není v tom takový problém, jako kdyby chtěl zbohatnout a pravidelně prohrával celou výplatu. Obzvlášť pokud má rodinu a veškeré peníze začne vkládat do sázek.

## SEZNAM LITERATURY

- ABZ.cz. (2021). *Predikce—ABZ.cz: Slovník cizích slov*. ABZ slovník cizích slov [online]. [cit. 2021-06-18]. Dostupné z: <https://slovník-cizich-slov.abz.cz/web.php/slovo/predikce>
- Associated Press. (2021). *Australian Open will use live electronic line calls*. ESPN.Com[online]. [cit. 2021-02-03]. Dostupné z: [https://www.espn.com.au/tennis/story/\\_/id/30824282/australian-open-courts-feature-live-electronic-line-calls-year](https://www.espn.com.au/tennis/story/_/id/30824282/australian-open-courts-feature-live-electronic-line-calls-year)
- ATP Tour. (2020). *Australian Open Set For Historic Start, Total Prize Pool Revealed | ATP Tour | Tennis*. ATP Tour [online]. [cit. 2020-12-19]. Dostupné z: <https://www.atptour.com/en/news/australian-open-press-release-19-december-2020>
- ATP Tour. (2021). *Rankings | Singles | ATP Tour | Tennis*. ATP Tour [online]. [cit. 2021-06-02]. Dostupné z: <https://www.atptour.com/en/rankings/singles/www.atptour.com/en/rankings/singles>
- ATP Tour. (2022). *Australian Open | Results | ATP Tour | Tennis*. ATP Tour [online]. [cit. 2022-11-13]. Dostupné z: <https://www.atptour.com/en/scores/archive/australian-open/580/2022/results?matchType=singles>
- ATP Tour, Inc. (2021). *The 2021 ATP® Official Rulebook*. ATP Tour, Inc. <https://www.atptour.com/en/corporate/rulebook>
- Barnett, T., & Clarke, S. R. (2005). Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*, 16(2), 113–120. <https://doi.org/10.1093/imaman/dpi001>
- BetArena.cz. (2015). *Počítání value sázky—Škola sázení | BetArena.cz*. Bet Arena [online]. [cit. 2021-05-20]. Dostupné z: [https://www.betarena.cz/rubriky/skola-sazeni/pocitani-value-sazky-skola-sazeni\\_9.html](https://www.betarena.cz/rubriky/skola-sazeni/pocitani-value-sazky-skola-sazeni_9.html)

- Bílý, J. (2003). *Právní dějiny na území ČR: vysokoškolská učebnice*. Linde.
- Bradley, R. A., & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, 39(3/4), 324–345.
- Brom, O. (2018). *Predikce | blog | ACREA CR spol. S r.o.* Acrea [online]. [cit. 2021-04-06].  
Dostupné z: <https://acrea.cz/predikce-nehadejte-z-karet-ale-pouzijte-statisticky-model/>
- Clarke, S. R., & Dyte, D. (2000). Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research*, 7(6), 585–594.  
<https://doi.org/10.1111/j.1475-3995.2000.tb00218.x>
- Corral, J., & Prieto-Rodríguez, J. (2010). Are differences in ranks good predictors for Grand Slam tennis matches? *International Journal of Forecasting*, 26(3), 551–563.  
<https://doi.org/10.1016/j.ijforecast.2009.12.006>
- Červinka, J. (2017). *Turnajové systémy*. Ostrov nápadů [online]. [cit. 2021-06-19]. Dostupné z: <https://ostrovnepadu.cz/turnajove-systemy/>
- Fačevicová, K., Hron, K., & Kunderová, P. (2018). *Markovovy řetězce a jejich aplikace (2.)*. Přírodovědecká fakulta - Univerzita Palackého v Olomouci.
- Grand Slam Board. (2020). *2020 OFFICIAL GRAND SLAM RULE BOOK*. Grand Slam® Board. <https://www.itftennis.com/media/2495/grand-slam-rulebook-2020-f.pdf>
- International Tennis Federation. (2019). *History of tennis courts*. International Tennis Federation [online]. [cit. 2021-06-18]. Dostupné z: <https://www.itftennis.com/media/2101/surfaces-history-of-tennis-courts.pdf>
- ITF LTD. (2021). *ITF RULES OF TENNIS*. ITF LTD.  
<https://www.itftennis.com/media/7221/2022-rules-of-tennis-english.pdf>
- Klaassen, F., & Magnus, J. R. (1998). On the Independence and Identical Distribution of Points in Tennis. *Econometrics*, 1998(53).

- Mastela, M., Dobešová, L., & Bajura, J. (2011). *Občanský zákoník v dotazech a odpovědích* (1.). Anag.
- Matchstat.com. (2022). *Head To Head Tennis Search | ATP & WTA H2H Search*. Matchstat [online]. [cit. 2022-11-10]. Dostupné z: <https://matchstat.com/tennis/head-to-head>
- McHale, I., & Morton, A. (2011). A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting*, 27(2), 619–630.  
<https://doi.org/10.1016/j.ijforecast.2010.04.004>
- Mitchell, C. (2022). *Fibonacci Sequence: Definition, How it Works, and How to Use It*. Investopedia [online]. [cit. 2022-08-16]. Dostupné z: <https://www.investopedia.com/terms/f/fibonaccilines.asp>
- Newton, P. K., & Keller, J. B. (2005). Probability of Winning at Tennis I. Theory and Data. *Studies in Applied Mathematics*, 114(3), 241–269. <https://doi.org/10.1111/j.0022-2526.2005.01547.x>
- peRFect Tennis. (2022). *Australian Open Prize Money 2022*. peRFect Tennis [online]. [cit. 2022-06-18]. Dostupné z: <https://www.perfect-tennis.com/prize-money/australian-open/>
- Petráčková, V., & Kraus, J. (2000). *Akademický slovník cizích slov: A-Ž*. Academia.
- Pinnacle. (2013). *Co je Fibonacciho systém sázení?* Pinnacle [online]. [cit. 2021-08-16]. Dostupné z: <http://www.pinnacle.com/cs/betting-articles/Betting-Strategy/what-is-the-fibonacci-betting-system/47LJRM4KLKLUZXXKX>
- Pollard, G. (2017). *A mathematical, statistical, probabilistic and strategic analysis of tennis* [Thesis, Swinburne University of Technology]. [https://researchbank.swinburne.edu.au/file/b693026c-94cf-46d5-9180-158972955b53/1/geoff\\_pollard\\_thesis.pdf](https://researchbank.swinburne.edu.au/file/b693026c-94cf-46d5-9180-158972955b53/1/geoff_pollard_thesis.pdf)

- Sazkovekancelare.cz. (2020). *Strategie sázení—Nejlepší strategie, které vám usnadní dosáhnout zisku*. Sázkové kanceláře [online]. [cit. 2021-05-23]. Dostupné z: <https://sazkovekancelare.cz/strategie-sazeni/>
- Simmonds, E., & O'Donoghue, P. (2018). Probabilistic models comparing Fast4 and traditional tennis. *International Journal of Computer Science in Sport*, 17(2), 141–162. <https://doi.org/10.2478/ijcss-2018-0008>
- Sipko, M., & Knottenbelt, W. (2015). *Machine learning for the prediction of professional tennis matches* [Master's thesis, Imperial College London]. <https://www.doc.ic.ac.uk/teaching/distinguished-projects/2015/m.sipko.pdf>
- Spanias, D. (2014). *Professional Tennis: Quantitative Models and Ranking Algorithms*. Imperial College London.
- Spanias, D., & Knottenbelt, W. J. (2013). Predicting the outcomes of tennis matches using a low-level point model. *IMA Journal of Management Mathematics*, 24(3), 311–320. <https://doi.org/doi.org/10.1093/imaman/dps010>
- Tennis Australia. (2021a). *Community heroes showcased as Live Electronic Line Calling introduced*. Australian Open [online]. [cit. 2021-05-29]. Dostupné z: <https://ausopen.com/articles/news/community-heroes-showcased-live-electronic-line-calling-introduced>
- Tennis Australia. (2021b). *Results | Australian Open*. Australian Open [online]. [cit. 2021-06-21]. Dostupné z: <https://ausopen.com/results#!All>
- Tennis Board. (2022). *Head to Head Comparison: Tennis H2H Predictions and Stats*. Tennisboard [online]. [cit. 2022-06-13]. Dostupné z: <https://www.tennisboard.com/headtohead>
- Trojan, O. (2021). *Sázka, hra, los*. Holec, Zuska & Partneři [online]. [cit. 2021-12-12]. Dostupné z: <https://www.holec-advokati.cz/cs/sazka-hra-los/>

US Open. (2022). *Results—Official Site of the 2022 US Open Tennis Championships*. US Open [online]. [cit. 2022-11-13]. Dostupné z:

[https://www.usopen.org/en\\_US/scores/index.html?promo=subnav](https://www.usopen.org/en_US/scores/index.html?promo=subnav)

Utathya, N. (2021). *Tennis rankings: How they work and difference between ATP and WTA systems*. Olympics [online]. [cit. 2021-05-06]. Dostupné z:

<https://olympics.com/en/featured-news/tennis-rankings-atp-wta-men-women-doubles-singles-system-grand-slam-olympics>

Vlček, M. (2015). *Odvážné smlouvy* [Bakalářská práce, Bankovní institut vysoká škola Praha]. [https://is.ambis.cz/th/paj9w/?zoomy\\_is=1](https://is.ambis.cz/th/paj9w/?zoomy_is=1).

Zákon č. 186/2016 Sb., zákon o hazardních hrách (2021).

<https://www.aspi.cz/products/lawText/1/86652/1/2>

## SEZNAM PŘÍLOH

Příloha č. 1: Statistiky prvních 4 kol Australian Open 2021 hráče: Andrey Rublev.

Příloha č. 2: *Statistiky prvních 4 kol Australian Open 2021 hráče: Rafael Nadal.*

Příloha č. 3: *Statistiky prvních 4 kol Australian Open 2021 hráče: Alexander Zverev*

Příloha č. 4: *Statistiky prvních 4 kol Australian Open 2021 hráče: Aslan Karatsev*

Příloha č. 5: *Statistiky prvních 4 kol Australian Open 2021 hráče: Stefanos Tsitsipas*

Příloha č. 6: *Statistiky prvních 4 kol Australian Open 2021 hráče: Daniil Medvedev*

Příloha č. 7: *Statistiky prvních 4 kol Australian Open 2021 hráče: Novak Djoković*

Příloha č. 8: *Statistiky prvních 4 kol Australian Open 2021 hráče: Dimitrov Grigor*