

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF INTELLIGENT SYSTEMS

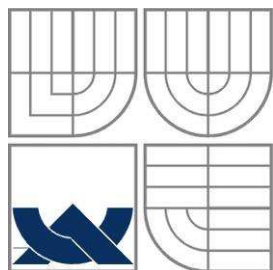
AGENT PRO KURZOVÉ SÁZENÍ

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

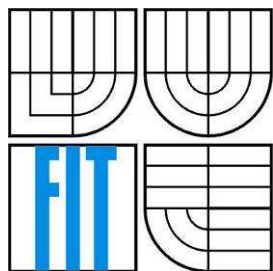
AUTOR PROJEKTU
AUTHOR

Bc. Jiří Bělohlávek

BRNO 2008



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF INTELLIGENT SYSTEMS

AGENT PRO KURZOVÉ SÁZENÍ
THE BETTING AGENT

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PROJEKTU
AUTHOR

Bc. Jiří Bělohlávek

VEDOUCÍ PROJEKTU
SUPERVISOR

Ing. Lukáš Grulich

BRNO 2008

Zadání

Originál zadání diplomové práce následuje bezprostředně za titulním listem.

Licenční smlouva – 1. STRANA

Vyplněná licenční smlouva dle přílohy rozhodnutí rektora č.9/2007. Licenční smlouvu obdrží student současně se zadáním diplomové práce.

Licenční smlouva – 2. STRANA

Vyplněná licenční smlouva dle přílohy rozhodnutí rektora č.9/2007. Licenční smlouvu obdrží student současně se zadáním diplomové práce.

Abstrakt

Tato práce vysvětluje návrh a implementaci agenta pro podporu kurzového sázení. Nechybí ani teoretické pozadí o online kurzovém sázení, pravděpodobnosti a statistice. V kapitole o dolování dat je nastíněn princip získávání znalostí z datových skladů a metody řešení různých typů úloh. Podrobněji je zpracována metoda řešení pomocí neuronových sítí např. pomocí algoritmu backpropagation. Najdete zde i množství grafů a histogramů, z analýzy dat, vytvořených programem SAS Enterprise Miner. V závěru práce se nachází shrnutí výsledků a další možnosti rozšíření agenta.

Klíčová slova

Dolování dat, datové sklady, metody dolování dat, neuronové sítě, model neuronu, matematický popis neuronu, topologie neuronové sítě, backpropagation, statistiky, pravděpodobnost, sportovní sázení, hazardní hra, online hazard, herní nabídka, sázky v přímém přenosu, strategie kurzového sázení, návratnost, jistá výhra, systémy asijských handicapů, systém 1-6 nula, sázení na hodnotu, vliv nedávné minulosti, potenciál a kapacita, analýza kurzů a zápasů, tipování, sázkové kanceláře, učení neuronové sítě, chyba neuronové sítě, databáze, Java, NetBeans IDE, SAS Enterprise Miner, grafy, histogramy, sport, tenis, MySQL.

Abstract

This master thesis deals with design and implementation of betting agent. It covers issues such as theoretical background of an online betting, probability and statistics. In its first part it is focused on data mining and explains the principle of knowledge mining from data warehouses and certain methods suitable for different types of tasks. Second, it is concerned with neural networks and algorithm of back-propagation. All the findings are demonstrated on and supported by graphs and histograms of data analysis, made via SAS Enterprise Miner program. In conclusion, the thesis summarizes all the results and offers specific methods of extension of the agent.

Keywords

Data mining, data warehouses, methods of knowledge mining, neural networks, model of neuron, mathematic description of neuron, topology of neural network, backpropagation, statistic, probability, sport betting, hazard game, online hazard, game offer, life betting, betting strategy, payback, sure bet, systems of Asian handicap, system 1-6 zero, valuebetting, recency effect, potential and capacity, odds and matches analysis, tipping, betting offices, learning of neural networks, error of neural network, database, Java, , NetBeans IDE, SAS Enterprise Miner, graphs, histograms, sport, tennis, MySQL.

Citace

Jiří Bělohlávek: Agent pro kurzové sázení, diplomová práce, FIT VUT v Brně, rok 2008.

Agent pro kurzové sázení

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením pana Ing. Lukáše Grulichy.

Další informace mi poskytli pan Ing. Jaroslav Rozman a Ph.D. Ing. Vladimír Bartík. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....
Jméno Příjmení

Datum

Poděkování

Děkuji svému vedoucímu práce panu Ing. Grulichovi za poskytnutou pomoc a cenné rady. Dále bych rád poděkoval za materiály, které mi zpřístupnil pan Ing. Jaroslav Rozman. Také děkuji svému kolegovi z projektu předmětu ZZN Bc. Vojtěchu Šmerdovi, který mi pomohl jít často do větší hloubky problému a vidět věci z více úhlů pohledu.

© Jiří Bělohlávek, 2008

Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.

Obsah

Obsah.....	8
1 Úvod.....	10
2 Použité technologie - teoretické koncepty.....	11
2.1 Dolování dat, hledání skrytých souvislostí.....	11
2.1.1 Definice dolování dat.....	11
2.1.2 Dolování dat a datové sklady.....	11
2.1.3 Typy úloh a metody pro jejich řešení.....	12
2.1.4 Metody dolování dat.....	12
2.2 Neuronové sítě.....	14
2.2.1 Model neuronu.....	14
2.2.2 Matematický popis neuronu.....	15
2.2.3 Umělá neuronová síť.....	16
2.2.4 Vrstvová struktura umělé neuronové sítě.....	16
2.2.5 Typy neuronových sítí.....	16
2.2.6 Backpropagation.....	17
2.3 Statistika.....	18
2.3.1 Matematická statistika.....	19
2.3.2 Teorie pravděpodobnosti.....	19
3 Sportovní sázení.....	21
3.1 Hazardní hra.....	21
3.2 Účast na online hazardu.....	21
3.3 Podmínky pro účast.....	21
3.4 Herní nabídka.....	22
3.5 Výhody sázení po internetu z pohledu hráče.....	22
3.5.1 Dostupnost a pohodlí.....	22
3.5.2 Sázky v přímém přenosu.....	22
3.5.3 Výhodnější sázení.....	22
3.6 Strategie kurzové sázení.....	23
3.6.1 Základní myšlenky a strategie.....	23
3.6.2 návratnost.....	23
3.6.3 Sure bet (jistá výhra).....	24
3.6.4 Systémy asijských handicapů.....	24
3.6.5 Systém 1 - 6 nula.....	26
3.6.6 Sázení na hodnotu, valuebetting.....	27

3.6.7	Jak sázet a vyhrávat	29
3.6.8	Recency Effect (Vliv nedávné minulosti).....	30
3.6.9	Potenciál a kapacita v kurzovém sázení.....	31
3.6.10	Handicapping - jak analyzovat hru nebo zápas.....	31
3.6.11	Jak tipovat a čeho se vyvarovat při sázení po internetu	32
3.7	Sázkové kanceláře	33
3.7.1	Jak si zvolit online sázkovou kancelář?	33
4	Návrh a struktura aplikace.....	35
4.1	Neuronová síť.....	35
4.1.1	Učení neuronové sítě.....	36
4.1.2	Chyba sítě a přesnost	37
4.2	Zdrojová data.....	38
4.2.1	Zdrojová data popis.....	39
4.2.2	Zdrojová data pro neuronovou síť	40
4.3	Databáze	40
4.3.1	Práce s databází v programovacím jazyku Java.....	42
4.4	Pomocné soubory	42
4.5	Použité nástroje	43
4.6	Další zdroje dat.....	43
5	Výsledky.....	44
5.1	Implementace v NetBeans IDE 6.0	44
5.2	SAS Enterprise Miner.....	49
5.2.1	Charakteristika proměnných	49
5.2.2	Grafy a histogramy	50
5.	Budoucí rozšíření.....	56
6	Závěr.....	58
	Literatura	59
	Seznam příloh.....	62
	Příloha 1, zdrojové texty.....	63

1 Úvod

Jedním z mnoha trendů naší doby se stalo kurzové sázení. Nejen, že existuje mnoho kamenných sázkových kanceláří, ale co je důležitější, sázet je možné i prostřednictvím virtuálních sázkových kanceláří přístupných online na internetu.

Internet je medium, na kterém nalezneme většinu dostupných informací a není tomu jinak i u sportovního sázení. Nemusíte si kupovat různé sportovní deníky, nebo časopisy, ale stačí se připojit na několik webových stránek a zjistíte vše potřebné. Nechybí statistiky klubů, hráčů, vzájemné bilance, historie zápasů, kritiky předchozích strategií trenérů, informace o zranění hráčů, událostech v klubech aj. Můžete si koupit, nebo zadarmo vyzkoušet, řadu programů pro podporu rozhodování při sportovním sázení, několikrát denně dostávat sportovní updaty, nebo si jen předplatit sportovní tipy. Sázkové kanceláře nás lákají zajímavými bonusy pro nově registrované a je také možné sázet na tzv. burzách sázek.

Tato práce si klade za cíl osvětlit problematiku využití internetu, výpočetní techniky a možností umělé inteligence pro kurzové sázení. Zabývá se problémy, jako jsou volba strategií, teoretické pozadí sázek, statistika, dolování znalostí z databází, neuronovými sítěmi a charakteristikou sportovního sázení.

Po prostudování výše zmíněných teoretických oblastí jsem implementoval agenta pro podporu kurzového sázení v jazyce Java. Agent je schopný stahovat zápasy z informačních portálů pro sportovní kurzové sázení. Pomocí neuronové sítě provádí odhady kurzů jednotlivých zápasů. Neuronová síť je trénována na základě výsledů zápasů za několik posledních let. Agent má online přístup k aktuálním žebříčkům hráčů, jejich zjednodušeným profilům, aj.

2 Použité technologie - teoretické koncepty

Kapitola popisuje nejdůležitější techniky a teoretické principy, ze kterých vychází analýza a implementační řešení agenta pro podporu kurzového sázení.

Následující podkapitoly pojednávají o dolování znalostí z databází, neuronových sítích a statistice. Základní myšlenkou je nahradit statistické vzorce pro výpočty kurzů na sportovní příležitosti neuronovou sítí. Tato síť bude trénována „informačně bohatými“ daty z databáze zápasů hraných za několik posledních let. Právě „informačně bohatá“ data nám pomohly vybrat techniky dolování.

2.1 Dolování dat, hledání skrytých souvislostí¹

2.1.1 Definice dolování dat

Obecně je přijata následující definice dolování dat, známá též jako získávání znalostí: *Dolování dat je proces výběru, prohledávání a modelování ve velkých objemech dat sloužící k odhalení dříve neznámých vztahů mezi daty za účelem získání obchodní výhody.*

Obchodní výhoda označuje cíl dolování - dolování by vždy mělo mít za cíl řešení konkrétního obchodního problému či nalezení cesty k vylepšení procesu. Cíl by měl být předem definován a jen na jeho základě by se měla připravovat data. Data pro dolování by ideálně měla být brána z datového skladu zahrnujícího historické hodnoty z různých podnikových systémů. Příprava dat je úzce svázána s pojmem proces. Ten znamená, že dolování není jednorázová analýza, ale že nasazení technologie dolování dat předpokládá přípravu podnikových procesů umožňujících kontinuální využívání analýz a podporujících zpětné vazby od uživatelů. Ty pak mohou ovlivňovat jak proces sběru dat, tak definice nových cílů.

2.1.2 Dolování dat a datové sklady

Existuje mnoho důvodů pro úzkou návaznost datového skladu a dolování dat. Nejdůležitějším důvodem je kvalita vstupních dat pro dolování. Sebedokonalejší modelovací technika či analýza nepřinesou očekávaný výsledek, pokud nejsou vstupní data očištěna od chyb, zkontrolována úplnost všech požadovaných údajů a sjednoceny formáty z různých systémů. Procesní charakter dolování dat vyžaduje, aby se jako vstup dolování dat používala průběžně aktualizována data. Všem těmto požadavkům vyhovuje datový sklad. Rozšíření datových skladů je naopak jedním z hybných prvků bouřlivého rozvoje technologií dolování dat.

¹ Viz [1]

2.1.3 Typy úloh a metody pro jejich řešení²

Existují jednak různé typy úloh, které je možno řešit, jednak různé postupy, které lze při řešení použít. Základní klasifikace je uvedena v následující tabulce.

Úloha	Metoda
Klasifikace	Diskriminační analýza Logistická regresní analýza Klasifikační (rozhodovací) stromy Neuronové sítě (algoritmus "backpropagation")
Odhady hodnot vysvětlované proměnné	Lineární regresní analýza Nelineární regresní analýza Neuronové sítě (RBF -- "radial basis function")
Segmentace (shlukování)	Shluková analýza Genetické algoritmy Neuronové shlukování (Kohonenovy mapy)
Analýza vztahů	Asociační algoritmus pro odvozování pravidel typu If X, then Y
Predikce v časových řadách	Boxova-Jenkinsova metodologie Neuronové sítě ("recurrent back propagation")
Detekce odchylek	Vizualizace Statistické postupy

2.1.4 Metody dolování dat

Dolování dat je označením třídy úloh, které řeší mnohdy zdánlivě nesouvisející problémy z nejrůznějších oborů. Je pozoruhodné, kolik praktických aplikací má několik obecných metod analýzy dat. Výběr metody, která bude použita pro řešení daného problému, je jen jedním z kroků procesu dolování dat. Je třeba mít na zřeteli cíl, pro jehož splnění lze použít více metod. Pak je dobré znát jejich výhody a mít možnost porovnat jejich výsledky.

2.1.4.1 Prediktivní modelování

Prediktivní modelování je postupem, kdy se na základě známé množiny vstupních a známých výstupních hodnot, jím odpovídajících, hledá nejpravděpodobnější hodnota výstupu pro předem neznámé kombinace vstupních hodnot.

Elementárním příkladem prediktivního modelování je např. hodnocení rizika úvěru v bankovníctví. Banka soustřeďuje mnoho záznamů o svých minulých klientech a ze zkušenosti ví, kteří jsou špatnými dlužníky. Po vytvoření prediktivního modelu popisujícího hodnocení dlužníka

² Viz [2]

(výstup) na základě informací o něm (vstupní data) lze ohodnocovat rizika nově příchozích zákazníků (o kterých lze zjistit údaje používané jako vstupní data modelu). Používanými technikami pro prediktivní modelování jsou různé typy regrese, neuronové sítě a rozhodovací stromy.

Regrese

Regrese je standardní statistická metoda schopná popsat stupeň důležitosti vstupních proměnných na výstupu. Její síla tkví v teoretické propracovanosti odhadu chyb modelu a možnosti hledat i závislosti na kombinaci vstupních proměnných. Dobře jsou propracovány metody regrese pro odhad binárního výstupu (logistická regrese) či regrese pracující s daty nabývajících jen diskrétních (i nečíselných) hodnot. Použití regrese je limitováno pracností a časovou náročností vývoje složitějších modelů.

Neuronové sítě

Neuronové sítě jsou novou moderní technikou prediktivního modelování vynikající velkou variabilitou možných modelů a snadností modifikace jejich návrhu. Z pohledu regrese jsou neuronové sítě elegantní technikou pro hledání parametrů modelu založeného na velice flexibilním systému vnořených funkcí. Na druhé straně model založený na neuronové síti nemá srozumitelnou interpretaci.

Rozhodovací stromy

Rozhodovací stromy naopak získaly popularitu díky své snadné interpretaci. Popis modelu pomocí rozhodovacího stromu je řadou jednoduchých rozhodovacích pravidel často prezentovaných ve formě grafu. Tyto grafy mohou být snadno, bez hlubokých znalostí statistických metod, interpretovány řídicími pracovníky. Při použití všech technik modelování je nutno řešit problémy s volbou počtu parametrů modelu, jejich konvergenci a odhad chyb.

2.1.4.2 Klasifikace

Obecně je klasifikace metodou pro rozdělování dat do skupin dle jistých kritérií. Pokud jsou tato kritéria předem známa, alespoň pro vzorek dat, lze pomocí metod prediktivního modelování vyvinout model, jehož výstupem je klasifikační proměnná. Mnohem častější případ je neřízená klasifikace, kdy výsledná kritéria nejsou předem známa a úlohou klasifikace je jejich nalezení. Používanou technikou v takovýchto případech je shluková analýza (Cluster Analysis). Elementárním příkladem shlukové analýzy je např. nalezení skupin obchodů na základě jejich obratu, sortimentů a typu zákazníků. Nalezené skupiny lze pak použít např. pro specifikaci reklamní kampaně zaměřené na jednotlivé skupiny prodejen.

2.1.4.3 Analýza asociací

Nejčastějším použitím analýzy asociací, a zároveň jejím ilustrativním příkladem, je tzv. analýza nákupního košíku. Ta se zabývá hledáním kombinací produktů, které se ve vstupních datech (nákupním koši spotřebitelů) vyskytují významně časněji spolu. Cílem je odhalit pravidla typu: při

nákupu zboží A a C spotřebitele výrazně častěji nakupují zboží D a B. Odhalení takovýchto kombinací pomáhá marketingovým odborníkům v organizování nabídky či společných balíčků produktů.

2.1.4.4 Vzorkování

Vzorkování není algoritmem řešící přímo nějaký zadaný úkol dolování dat, ale je to jedna ze základních technik dolování dat umožňující získat výsledek v rozumném čase. Vzorkování je výběr omezené množiny dat ze základního souboru. Nejjednodušším způsobem vzorkování je náhodný výběr, který slouží jen ke zmenšení objemu zpracovávaných dat a tím k zrychlení výpočtů. Složitější metody vzorkování, např. výběr stejného počtu záznamů daného typu (stejný počet záznamů z každého tarifního pásma), umožňují redukci objemu zpracovávaných dat při současné záruce požadované přesnosti výsledku.

Vzorkování databáze pro analýzy je jistě neobvyklé z pohledu klasických databázových aplikací užívajících SQL nástrojů, ale objevují se již úvahy o nutnosti zařazení takovýchto služeb do standardních databázových serverů.

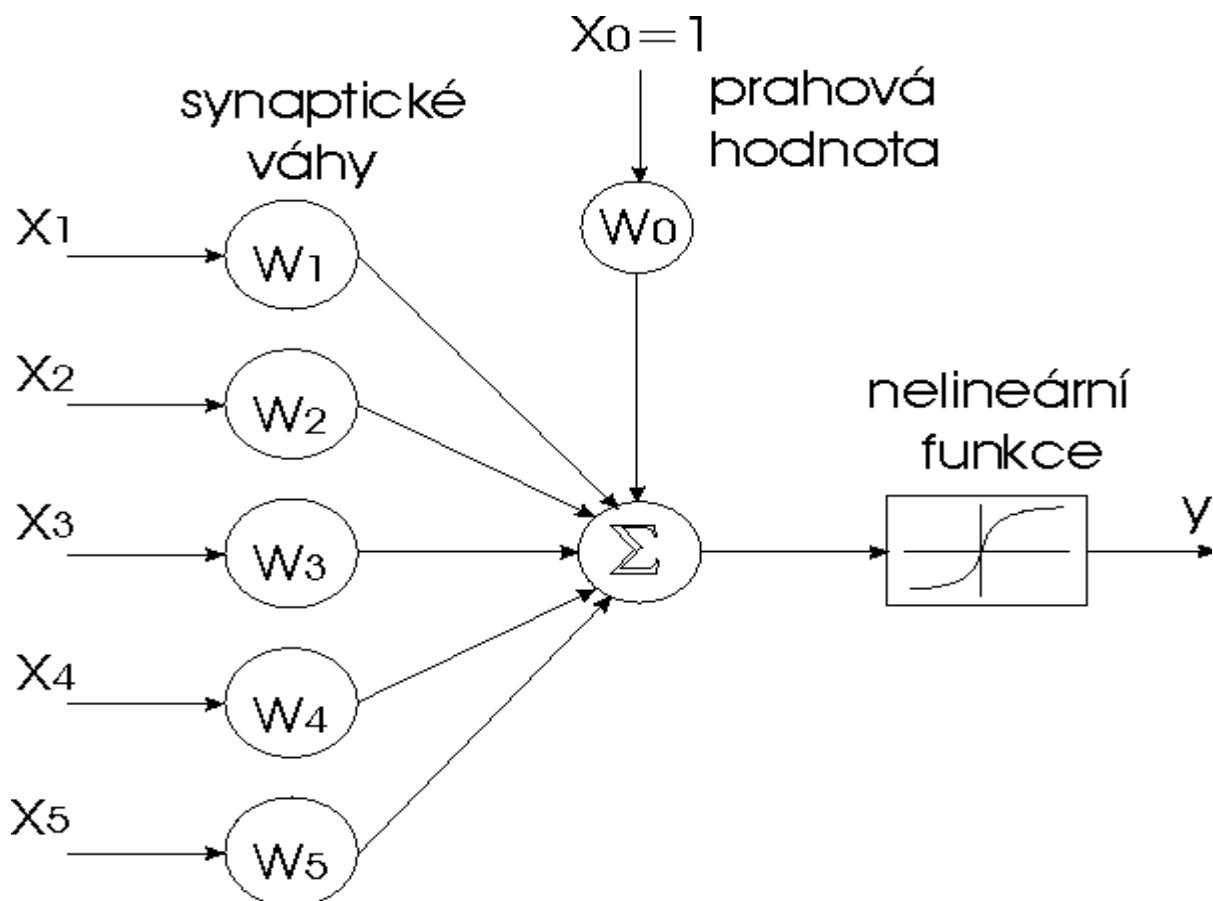
2.2 Neuronové sítě

Umělé neuronové sítě byly vytvořeny na základě jednoduchých modelů neuronů - funkčních buněk nervového systému živých organismů. Většina současných aplikací umělých neuronových sítí využívá selektivní a generalizační vlastnosti těchto struktur. Některé novější struktury jsou navíc schopné řešit i úlohy složitějšího typu, jako jsou např. optimalizační úlohy.

2.2.1 Model neuronu³

První matematický model neuronu vytvořili McCulloch a Pitts v roce 1943 a tento model se dodnes používá pro běžné aplikace. Tento matematický model se skládá ze tří hlavních částí. Obsahuje vstupní, výstupní a funkční část. Vstupní část se skládá ze vstupů a z přiřazených, nastavitelných vah (synaptické váhy). Na základě váhových koeficientů mohou být jednotlivé vstupy zvýhodňovány či potlačeny. Následující částí je výkonná jednotka, která zpracuje informace ze vstupu a vygeneruje výstupní odezvu. Třetí část je výstupní jednotka, která přivádí výstupní informace na vstup jiných neuronů. Z toho je patrná podoba mezi klasickými výpočetními systémy a umělými neurony. Oba systémy obsahují vstupní část, paměť, výkonnou jednotku a výstupní část. Velké rozdíly jsou ovšem v uspořádání těchto částí. Paměť umělého neuronu není samostatná jednotka, ale je rozprostřena ve vstupní části formou váhových koeficientů. Pomocí těchto koeficientů je systém schopný zapamatovat si informace. Jak je vidět na obrázku obr. 1, výkonná jednotka umělého neuronu je mnohem jednodušší než výkonná jednotka výpočetních systémů a je tvořena jednoduchou nelineární funkcí.

³ Viz [3]



Obrázek 1: Jednoduchý model neuronu

Z obrázku obr. 1 je jasná funkce jednoho neuronu. Vstupní hodnoty jsou vynásobeny příslušnými váhovými koeficienty a sečtou se. Na výsledek součtu se aplikuje funkce (obecně nelineární) a výsledná hodnota funkce je přivedena na vstup jiných neuronů pomocí výstupní části. Na obr. 1 je navíc vidět, že neuron má jeden zvláštní vstup, který není připojený k výstupu žádného neuronu, ale přivádí konstantní veličinu do neuronu. Tato veličina funguje jako prahová hodnota při aktivování výstupu. Když suma váženého součtu vstupů nepřesahuje prahovou hodnotu, tak se neuron neaktivuje a jeho výstup zůstane nezměněný.

2.2.2 Matematický popis neuronu

Matematicky lze funkci neuronu popsat následovně:

kde:

- x_i - je hodnota na i -tém vstupu,
- w_i - je váha i -tého vstupu,
- Q - je prahová hodnota,
- n - je celkový počet vstupů,
- F - je obecná nelineární funkce,
- y - je hodnota výstupu.

$$y = F\left(\sum_{i=1}^n x_i w_i + \Theta\right)$$

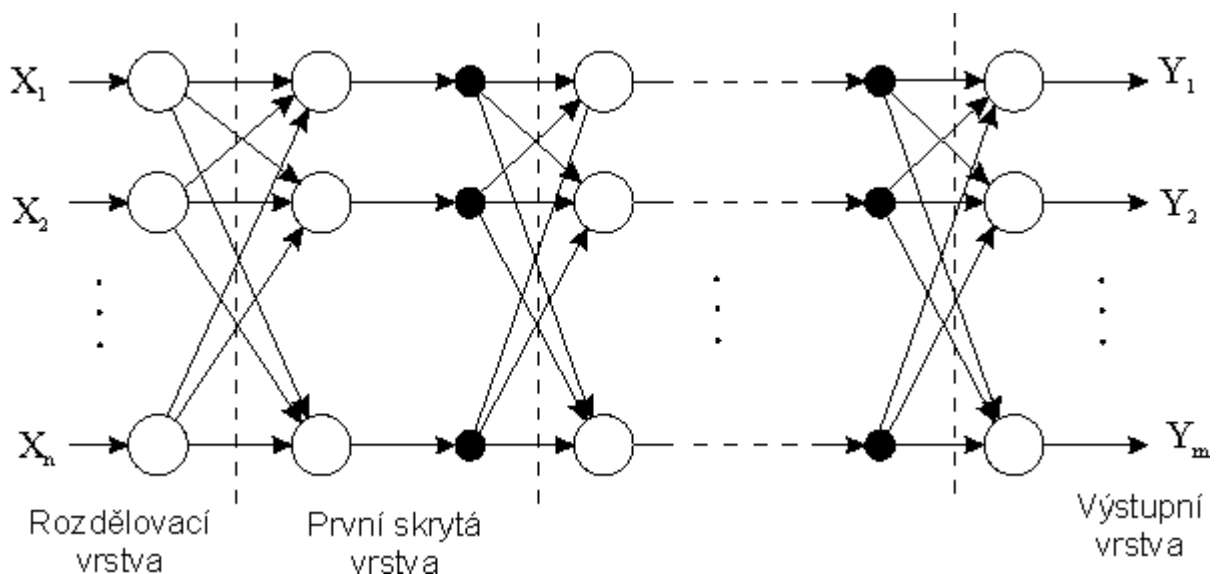
2.2.3 Umělá neuronová síť

Je asi zřejmé, že jediný neuron není schopen vykonat příliš složitou funkci. Síla systému, využívající umělé neurony, je ve struktuře, v síti velkého počtu neuronů. Umělá neuronová síť je vlastně pole jednoduchých výkonných prvků - neuronů. Takovéto uspořádání má velkou flexibilitu a spolehlivost. Umožňuje různě propojovat vstupy a výstupy neuronů, zvýhodnit či potlačit některé vstupy a minimalizovat vliv nesprávně fungujícího neuronu na celkový výsledek.

Samozřejmě i tento systém má nevýhody. Největší problémy se vyskytují při realizaci velmi složitých struktur, kde velký počet propojení mezi neurony se realizuje velmi obtížně. Dalším problémem je, že neexistuje jednoznačný postup při syntéze složitějších struktur.

2.2.4 Vrstvová struktura umělé neuronové sítě

Neurony jsou většinou sdružovány do vrstev, jak to ukazuje obr. 2. Výstupy z n -té vrstvy jsou přivedeny na vstup obecně každého neuronu ve vrstvě $n+1$. První vrstva se nazývá vstupní či rozdělovací a má za úkol přijímat hodnoty z okolí pro zpracování a přivést je na vstup každého neuronu následující vrstvy. Poslední vrstva nese název výstupní a hodnoty na jejím výstupu jsou odezvou celého systému na vstupní vzorky. Vnitřní vrstvy se nazývají skryté vrstvy. Jejich počet závisí na složitosti funkce, kterou má síť vykonat a na zvoleném typu sítě.



Obrázek 2: Vrstvová struktura umělé neuronové sítě

2.2.5 Typy neuronových sítí

Neuronové sítě lze rozdělit do dvou hlavních skupin podle struktury: na sítě s dopředním šířením signálu a na sítě se zpětnou vazbou. V současnosti se nejčastěji používají struktury s dopředním šířením signálu, kde výstupy z jedné vrstvy jsou vedeny na vstup následující vrstvy, jak to ukazuje obr. 2. Výstupy z poslední vrstvy jsou výstupy z celé sítě.

Struktura sítí se zpětnou vazbou se liší od předchozích v tom, že výstupy z vrstvy jsou vedeny zpět na vstupy dané vrstvy. Taková struktura umožňuje realizovat výpočty založené na iteračním procesu, a tak řešit např. optimalizační úlohy.

Neuronové sítě s dopředním šířením signálu lze rozdělit do dvou skupin podle funkce kterou realizují, a to na lineární a nelineární. Tato funkce samozřejmě není totožná s výstupní funkcí jednoho neuronu. Sítě lineární jsou schopné realizovat lineární matematické funkce, tj. funkce skládající se ze součtů a z násobení. Jednoduchý příklad lineární funkce může být logická funkce AND realizovaná pomocí jednoho neuronu. Charakteristickou vlastností nelineárních neuronových sítí s dopředním šířením signálu je schopnost učení. Fáze učení předchází fázi vlastní práce a slouží k určení váhových koeficientů, a tak vlastně k uložení informací do paměti systému.

Učení může probíhat dvěma způsoby, s učitelem a bez učitele. Při prvním způsobu je síť trénována pomocí dvojic vstupní vzorek a příslušný, očekávaný výstupní vzorek. Trénovací vstupní vzorky jsou vybrány z celkové množiny vstupních vzorků tak, aby plně popsaly všechny vlastnosti množiny důležité pro danou úlohu. V této fázi nenatrénované síti přiložíme vstupní vzorek. Na základě skutečné odezvy a očekávané odezvy se upravují váhové koeficienty. Během trénování se na vstupy sítě přivedou všechny trénovací vzorky, obecně vícekrát a navíc v náhodném pořadí. Po natrénování musí síť správně reagovat na všechny trénovací vzorky a dále má pracovat dobře i pro ostatní vzorky množiny. Aby síť pracovala dobře, potřebujeme velký počet trénovacích vzorků. Obecně platí, že čím větší je počet trénovacích vzorků, tím přesněji bude síť pracovat. Příkladem takové sítě je síť **"backpropagation"**, která je pravděpodobně nejčastěji používaným typem.

Při učení bez učitele máme jenom trénovací vzorky, ale neexistují očekávané výstupní vzorky. Tyto výstupní vzorky, příslušející k jednotlivým vstupním vzorkům, se určí během procesu učení. Váhové koeficienty se postupně nakonfigurují tak, aby pro každý vstupní trénovací vzorek existoval jediný aktivní výstup. Tak na konci trénování dosáhneme toho, že přivedením trénovacího vzorku se aktivuje vždy jediný, jednoznačně určený výstup.

2.2.6 Backpropagation⁴

Backpropagation je algoritmus, který byl vytvořen pro učení vícevrstevných neuronových sítí s učitelem. Tento algoritmus opravuje-nastavuje váhy jednotlivých spojů zpětným chodem tak, aby jejich velikosti byly z hlediska řešeného problému pokud možno optimální - hledá se globální minimum chybové funkce. Nastavení vah tedy probíhá v opačném směru, než jakým se šíří vstupní informace. U tohoto algoritmu rozeznáváme dvě fáze - aktivační a adaptační. Tato metoda je vlastně zobecněným delta pravidlem pro vícevrstevnou perceptronovou síť.

⁴ Viz [4]

2.2.6.1 Aktivační fáze.

Při inicializaci sítě se musí nastavit váhy na vhodnou hodnotu, což se dělá obvykle pomocí generátoru náhodných čísel. Váhy se pohybují v rozmezí 0,5 až -0,5. Kromě použití generátoru náhodných čísel existují i jiné metody pro prvopočáteční nastavení vah jako simulované žíhání a genetické algoritmy. Aktivační fáze je používána při učení a vybavování sítě. Je to tedy aktivita, při které se vstupní informace dostane na výstup a je zmodifikována momentální množinou vah a přenosovými funkcemi ve vlastních neuronech. Základní algoritmus, pomocí kterého se příslušná vícevrstvá neuronová síť může učit je backpropagation. Někdy bývá v literatuře síť využívající tento algoritmus chybně nazývána jako "backpropagation síť".

Backpropagation je pouze algoritmus, který byl vytvořen pro učení vícevrstevných neuronových sítí s učitelem. Tento algoritmus opravuje-nastavuje váhy jednotlivých spojů zpětným chodem tak, aby jejich velikosti byly z hlediska řešeného problému pokud možno optimální - hledá se globální minimum chybové funkce. Nastavení vah tedy probíhá v opačném směru, než jakým se šíří vstupní informace. U tohoto algoritmu opět rozeznáváme dvě fáze - aktivační a adaptační.

2.2.6.2 Adaptační fáze (backpropagation).

Výstupní vektor (odezva sítě na vstupní vektor) je porovnán s požadovaným originálem a rozdíl mezi oběma vektory je použit pro výpočet nových vah tak, že se nejprve opraví váhy u spojů, které vstupují do výstupní (nejvyšší) vrstvy. Pak jsou opraveny váhy u nižší vrstvy, atd. Až se dosáhne vrstvy vstupní, je tato fáze ukončena a opakuje se fáze aktivační. Při každém porovnávání výstupní odezvy s požadovaným originálem se daný rozdíl uchová v paměťové proměnné a sumarizuje se s dalšími postupně získanými rozdíly. Takto získané číslo za celou trénovací množinu (epochu) se nazývá globální chyba. Tato globální chyba je po každé epoše kontrolována s chybou, kterou zadal uživatel, a pokud je nižší než chyba zadaná, pak je síť naučena a proces učení končí. Algoritmus tedy hledá globální minimum chybové funkce.

2.3 Statistika⁵

Statistika je věda a postup jak rozvíjet lidské znalosti použitím empirických dat. Je založena na matematické statistice, která je větví aplikované matematiky. V teorii statistiky jsou náhodnost a neurčitost modelovány pomocí teorie pravděpodobnosti. Do praxe statistiky patří plánování, sumarizace a analýza nepřesných pozorování. Cílem statistiky je najít „nejlepší“ informace z dostupných dat, proto ji někteří autoři označují jako součást teorie rozhodování.

Jako statistiku také označujeme hodnoty, které získáme provedením náhodného výběru.

⁵ Viz [5]

Statistické postupy lze zhruba rozdělit na metody konfirmační analýzy, KA (např. intervaly spolehlivosti, regresní analýza ap.) a explorační analýzy, EA (např. shluková analýza, faktorová analýza, metoda hlavních komponent, metoda GUHA (kombinační analýza dat) apod.).

Základní rozdíl těchto dvou přístupů lze charakterizovat takto:

- V KA formulujeme hypotézu a metody KA použijeme k tomu, abychom ji potvrdili či vyvrátili.
- V EA máme k dispozici množství dat a od EA požadujeme, aby nám z nich vygenerovala nějaké hypotézy.

2.3.1 Matematická statistika

Matematická statistika tvoří teoretickou část statistiky. Na rozdíl od popisné statistiky či oficiální statistiky, se jedná o matematickou vědu aplikovanou přímo na problémy spojené se sběrem a pozorováním náhodných dat.

Statistika využívá metody, které jsou založeny na předpokladu, že zjišťované údaje jsou realizací náhodných veličin. Účelem shromažďování těchto dat je určení nějakého neznámého parametru jejich rozdělovacího zákona.

Statistika je úzce propojena s teorií pravděpodobnosti, která je dnes základním teoretickým oborem pro matematickou statistiku. V teorii pravděpodobnosti však na základě znalosti chování určité náhodné veličiny (tedy na základě zákona rozdělení pravděpodobnosti) určujeme pravděpodobnost určitého chování systému v dané situaci (tzn., určujeme výsledek náhodného pokusu). V matematické statistice je tomu naopak, tzn. na základě určitých dat (tzv. náhodného výběru) hledáme obecný model chování náhodné veličiny (obvykle charakteristiky rozdělení). Tento postup, tzn. hledání rozdělovacího zákona nebo charakteristik rozdělení, se označuje jako statistická indukce.

Základními typy statistické indukce jsou bodový odhad, intervalový odhad a testování hypotéz.

Statistickým modelem je nazýváno rozdělení pravděpodobnosti, na jehož základě získáváme číselné údaje pro statistický rozbor.

2.3.2 Teorie pravděpodobnosti

Teorie pravděpodobnosti (počet pravděpodobnosti) je matematická disciplína popisující zákonitosti týkající se jevů, které (přinejmenším z hlediska pozorovatele) mohou a nemusí nastat, resp., jejichž výsledná hodnota není předem jistá. Příkladem může být výsledek hodu kostkou ještě předtím, než hodíme, anebo venkovní teplota zítra v poledne. Takové jevy označujeme jako náhodné

2.3.2.1 Použití

Teorii pravděpodobnosti používáme v případech, kdy zkoumáme tzv. náhodné pokusy. Při náhodném pokusu není výsledek pokusu jednoznačně určen jeho počátečními podmínkami. To náhodné pokusy odlišuje od pokusů deterministických, jejichž výsledek je možno na základě určených počátečních podmínek jednoznačně určit.

Jsou-li tedy určeny počáteční podmínky pokusu, pak výsledek deterministického pokusu lze dopředu určit a při opakování pokusu (za stejných podmínek) bude výsledek deterministického pokusu stále stejný. Naproti tomu získáme při opakování náhodného pokusu různé výsledky (při stejných počátečních podmínkách), přičemž výsledek libovolného z těchto pokusů nelze předpovědět jednoznačně.

Náhodnost určitého pokusu je obvykle spojena s nedostatečnou znalostí počátečních podmínek daného pokusu. Kdybychom např. při hození kostkou byli schopni přesně určit všechny počáteční podmínky (poloha a orientace kostky v prostoru, její rychlost apod.), bylo by možné předpovědět, které číslo na kostce padne. Vzhledem k tomu, že tyto údaje neznáme, používáme k určení předpovědi metod teorie pravděpodobnosti.

Výsledků teorie pravděpodobnosti využívá zejména matematická statistika, zejména v oblasti asymptotického chování náhodných výběrů. Časté jsou také aplikace náhodných procesů na finanční, fyzikální a jiné procesy sledované v čase.

Dnes je teorie pravděpodobnosti široká disciplína zahrnující mnoho podoborů.

3 Sportovní sázení⁶

Sázení po internetu je elektronickou alternativou pro sportovní (kurzové) sázení a hazardní hry provozované v hernách a kasinech.

3.1 Hazardní hra

Společným jmenovatelem pro hazardní hry je kredit, reprezentující finanční vklad v určité hodnotě, který hráč musí do hry vložit, aby se jí mohl zúčastnit. Tento vklad se během hry rozděluje mezi hráče a provozovatele hry v závislosti na náhodě a způsobu hry. Pravidlem pro hazardní hry je, že z matematického hlediska je zisk ze hry vždy na straně provozovatele. Oproti tomu je v dlouhodobém horizontu hráč ve ztrátě, která vyplývá z podstaty hry. Nejatraktivnějším prvkem hazardních her je možnost rychlého zisku a adrenalin z rizika hry.

3.2 Účast na online hazardu

Ve většině případů provozují společnosti profitující z online hazardu jak internetovou sázkovou kancelář zaměřenou na kurzové sázení, tak kasino nabízející online verze her známých z kamenných kasin. Princip her a jejich provedení kopíruje jejich sesterské verze ze skutečného světa. Pro hráče, který se již v nějaké formě seznámil s některou z nabízených her, by neměl mít problém s online alternativou. Při kurzovém sázení se vyplňuje tiket podobně jako v offline sázkové kanceláři, ruleta disponuje hlasem (a někdy i zobrazením) krupiéra, v některých vícečlenných hrách mohou být simulováni protihráči. Stejně tak se jedné online hry může účastnit více hráčů bez ohledu na jejich momentální fyzickou polohu.

3.3 Podmínky pro účast

Rozhodujícím faktorem pro účast v hazardních hrách je věk. Spodní hranice věku hráče je daná zákony země, jejíž je hráč občanem. Ta je nejčastěji stanovena na 18, nebo 21 let.

Další podmínkou je finanční vklad. Online kasina často nabízejí hru s virtuálním vkladem, kdy hráč neriskuje vlastní hotovost, ale hraje nanečisto. Nejedná se tedy o hazard v definovaném slova smyslu. Pro hru s finanční hotovostí je nutné na účet provozovatele online kasina/herny složit finanční vklad, který se potom promítne do hráčova herního účtu. Podle technických možností provozovatele je možné se vkladem volně manipulovat a využívat je při účasti ve všech nabízených hrách. Stejně tak je možné nechat si jej zaslat nazpět.

⁶ Viz [6]

Třetí podmínkou je identifikace hráče, která slouží k prevenci hraní nezletilých. Dále je nutné dostačující technické vybavení (počítač, internetové připojení), a přiměřená znalost hry.

3.4 Herní nabídka

Herní nabídka se liší podle zvolené sázkové kanceláře/kasina. Většina provozovatelů rozděluje svou nabídku na dvě hlavní části: sázková kancelář a kasino. Hry se liší nejen vzhledem a provedením, ale často i pravidly. Některé hry jsou rozšířené o funkce, které při „kamenných“ hrách nejsou k dispozici. Nejčastěji jde o statistiky, nebo historie tahů.

Typické rozdělení herní nabídky ve většině online sázkových kanceláří/kasin

- Kurzové sázky na sportovní, politické a společenské události
- Stolní hry (Ruleta, Craps)
- Karetní hry (Blackjack, Poker)
- Výherní automaty
- Video poker

3.5 Výhody sázení po internetu z pohledu hráče

Ve srovnání se sázením v kamenných hernách a kasinech má sázení po internetu následující výhody:

3.5.1 Dostupnost a pohodlí

Hráč se ke svému hernímu účtu může připojit prakticky z jakéhokoliv počítače s dostupností internetu. Při sportovním sázení bývají na vybrané zápasy k dispozici i stovky možných sázek zohledňující všechny aspekty hry, tedy nejen obvyklý výsledek hry (počet gólů, délka prodloužení, počet vhazování, počet rohů atd. např. ve fotbale).

3.5.2 Sázky v přímém přenosu

Mnoho internetových sázkových kanceláří umožňuje sázení v průběhu zápasu. Kurzy na jednotlivé události v utkání se mění v závislosti na jeho průběhu.

3.5.3 Výhodnější sázení

Díky tomu, že si zákazník může vybrat z více sázkových kanceláří, vzniká o něj mezi provozovateli online hazardu přirozený boj. Jeho výsledkem jsou u kurzového sázení výhody ve formě konkurujících si kurzů a prakticky absence manipulačního poplatku, jak jej známe z většiny „kamenných sázkových kanceláří“. U online kasin je to pak širší nabídka, jak v počtu her, tak v jejich finančních rovinách.

3.6 Strategie kurzové sázení

3.6.1 Základní myšlenky a strategie⁷

Krásou sportovního sázení oproti kasinům je také v tom, že nelze přesně určit pravděpodobnost určitých situací. Práce bookmakera je se co nejvíce přiblížit a co nejlépe určit tyto pravděpodobnosti, většinou na základě různých statistik. Podle těchto pravděpodobností vypíše kurzy (návrtnost - pravděpodobnost). Další jeho prací je úprava kurzů podle náběru na určité příležitosti. Pokud sázkaři sází nejvíce na jednu příležitost (většinou na favorita), kurzy na něj bookmaker sníží, naopak kurzy na ostatní příležitosti zvýší (většinou tak, že zůstane stejná návratnost kurzů).

Zde začíná práce úspěšného sázkaře. Naprostým základem je umět to samé co bookmaker - určit pravděpodobnosti jednotlivých příležitostí. Na jejich základě zjistíme vlastní poctivé kurzy. Poté porovnáme nabídky kurzů na trhu. Pokud najdeme na nějakou příležitost vyšší kurz, než je náš poctivý kurz, našli jsme hodnotný kurz.

Určit tyto pravděpodobnosti ovšem není nic jednoduchého. Musíte brát v potaz statistiky - aktuální forma, výsledky doma/venku, vzájemné zápasy apod. Taktéž musíme zohlednit další faktory - zranění hráči (či naopak návraty zraněných hráčů) na jedné či druhé straně (a jejich důležitost pro tým), potřebnost bodů, únava hráčů, ale i počasí apod. Pokud s tím začínáte, bude dlouho trvat, než si tuto věc osvojíte, než si vytvoříte vlastní postupy. Nejlépe zjistíte jak se zlepšujete na základě statistik hypotetických sázek na Vámi nalezené hodnotové kurzy.

3.6.2 Návratnost⁸

Pojem návratnost vysvětlím na příkladě hodu mincí. Máme stejnou pravděpodobnost 50%, že nám padne pana nebo orel. Poctivý kurz na tuto pravděpodobnost odpovídá 2. Bookmakeři jsou ale nuceni vypsat kurz nižší, aby se užívali. To jak je kurz nízko ovlivňuje zdravá konkurence.

Přestavme si, že bookmaker vypíše kurz 1,8% na obě události při hodu mincí. Pokud si vsadíte například na to, že padne orel 1Kč, tak při výhře dostane 1,8Kč a návratnost je rovna podílu poctivého kurzu s nabízeným. Kvalitní sázkové kanceláře se snaží, aby návratnost byla co nejvyšší, obvyklá hodnota bývá okolo 90%.

3.6.2.1 Několik poznámek k návratnosti kurzů (z hlediska sportovního sázení):

- Čím je počet možných výsledků dané příležitosti vyšší, tím je obvykle návratnost nižší (kurzy na sázku typu 12 mají obvykle vyšší návratnost než kurzy na přesný výsledek).

⁷ Viz [7]

⁸ Viz [7]

- Čím je daná příležitost prestižnější, tím je návratnost obvykle vyšší. Návratnost kurzů na finále US Open je obvykle vyšší než návratnost na čtvrtfinále challengeru (v rámci jedné sázkové kanceláře-sk). Je to způsobeno konkurencí mezi jednotlivými sk (proto píšou obvykle, protože v českých sk toto pravidlo obvykle neplatí).
- Návratnost u českých sk na všechny příležitosti je kolem 90%. Návratnost u internetových sk značně kolísá podle typu sportu a specializace jednotlivých sk.
- Návratnost u některých příležitostech nelze jednoznačně určit (handicapy a under/over sázky s hladinou $\pm n$, kde n je celé číslo. Handicapy a under/over sázky s kombinovanou hladinou) - obecně u takových, kde se při určitém výsledku sázka vrací.

3.6.2.2 Návratnost a sázka na multiplikátor

Jestliže máte na vašem tiketu více zápasů, musíte si uvědomit, že se návratnosti mezi sebou násobí. Takže pokud máme akumulátor s 5 příležitostmi, dvě mají návratnost 90%, dvě 93% a jedna 94%, návratnost sázky na tento akumulátor je $0.92 \times 0.932 \times 0.94 = 65.85\%$, což je stejné jako bych Vám vypsal kurz 1.32 na pannu nebo orla.

3.6.3 Sure bet (jistá výhra)⁹

Princip sure bet je v tom, že máte účet u několika sázkových kanceláří a vyhledáváte kurzy v různých sázkových kancelářích, které se liší, a vy hledáte kurzy, které když vsadíte v určitém poměru, tak ať dopadne zápas jakkoliv, tak vyhraje vždycky. Možná si myslíte, že se kurzy u sázkových kanceláří moc neliší, ale opak je pravdou a pečlivým prohledáváním postupně sure bet najdete. Nejčastěji sure bet najdete při vypisování kurzů, kdy ještě sázkové kanceláře nesjednotí kurzy a dále při nějaké závažné situaci, která ovlivní kurz sázky, kdy některé sázkové kanceláře pomalu reagují při změně kurzu. Samozřejmě se sure bet dá najít i průběžně, převážně u sázkových příležitostech, které nejsou tak in a jejich kurzy se mohou diametrálně lišit.

3.6.4 Systémy asijských handicapů

Abychom pochopili silné stránky jednotlivých systémů, je vhodné si vytvořit tabulku možných výsledků a vypsat si, co to pro nás znamená u každého výsledku.

3.6.4.1 Asijský handicap 0 : 1¹⁰

Zvolíme ukázkový příklad, zápas Chelsea – Everton, Handicap: 0 : 1. Sázíte-li na výhru Chelsea: v případě jejího vítězství o dva góly vyhráváte svou sázku, zvítězí-li o 1 branku, dostanete zpět celý vklad v případě remízy či prohry vklad ztrácíte. Sázíte-li na výhru Evertonu: v případě jeho vítězství či

⁹ Viz [17]

¹⁰ Viz [18]

remízy vyhráváte svou sázku, v případě prohry Evertonu o 1 branku dostanete zpět celý vklad v případě prohry Evertonu o více, než 1 gól vklad ztrácíte.

Domáci	Hosté	Handicap	Kurs na Chelsea	Kurs na Everton
Chelsea	Everton	0 : 1	2.40	1.7

Příklad: Sázíte 100 USD na Everton

Výsledek	Chelsea	Everton	Poznámka
	0	1	Evertonu byla dán 1 gól k dobru (hned ze startu)
0:0	-100%	+100%	Everton vyhrává 1 : 0, získáváte vklad x kurs, tedy 170\$
1:0			Za této situace se sázky vracejí, nevyhráváte, neztrácíte
0:1	-100%	+100%	Everton vyhrává 2 : 0, získáváte vklad x kurs, tedy 170\$
2:0	+100%	-100%	Chelsea vyhrává 2 : 1, ztrácíte 100\$
3:0	+100%	-100%	Chelsea vyhrává 3 : 1, ztrácíte 100\$
1-1, 2-2	-100%	+100%	Everton vyhrává o 1 gól, získáváte vklad x kurs, tedy 170\$

3.6.4.2 Asijský handicap 0 : $\frac{3}{4}$ ¹¹

Zvolíme ukázkový příklad, zápas Chelsea – Everton, Handicap: 0 : $\frac{3}{4}$. Sázíte-li na výhru Chelsea: v případě jejího vítězství o dva góly vyhráváte svou sázku, zvítězí-li o 1 branku, dostanete zpět polovinu vkladu (ta druhá vyhrává) v případě remízy či prohry vklad ztrácíte. Sázíte-li na výhru Evertonu: v případě jeho vítězství či remízy vyhráváte svou sázku, v případě prohry Evertonu o 1 branku dostanete zpět polovinu vkladu (druhá půlka vyhrává) v případě prohry Evertonu o více než 1 gól vklad ztrácíte.

Domáci	Hosté	Handicap	Kurs na Chelsea	Kurs na Everton
Chelsea	Everton	0 : $\frac{3}{4}$	2.10	1.7

3.6.4.3 Asijský handicap 0 : $\frac{1}{2}$ ¹²

Zvolíme ukázkový příklad, zápas Chelsea – Everton, Handicap: 0 : $\frac{1}{2}$. Sázíte-li na výhru Chelsea: v případě jejího vítězství vyhráváte svou sázku, v případě remízy či prohry vklad ztrácíte.

Sázíte-li na výhru Evertonu: v případě jeho vítězství či remízy vyhráváte svou sázku, v případě prohry Evertonu vklad ztrácíte.

Domáci	Hosté	Handicap	Kurs na Chelsea	Kurs na Everton
Chelsea	Everton	0 : $\frac{1}{2}$	2.10	1.85

¹¹ Viz [19]

¹² Viz [20]

3.6.4.4 Asijský handicap 0 : ¼¹³

Zvolíme ukázkový příklad, zápas Chelsea – Everton, Handicap: 0 : ¼. Sázíte-li na výhru Chelsea: v případě jejího vítězství vyhráváte svou sázku, v případě remízy dostanete zpět polovinu vkladu. Sázíte-li na výhru Evertonu: v případě jeho vítězství vyhráváte svou sázku, v případě remízy dostanete zpět polovinu vkladu.

Domácí	Hosté	Handicap	Kurs na Chelsea	Kurs na Everton
Chelsea	Everton	0 : 1/4	1.95	2.0

3.6.4.5 Asijský handicap 0 : 0¹⁴

Zvolíme ukázkový příklad, zápas Chelsea – Everton, Handicap: 0 : 0. Pokud zápas skončí remízou, bude sázka vrácena v plné výši. Pokud trefíte vítěze, vyhráváte sázku, pokud jej netrefíte, svůj vklad ztrácíte.

Domácí	Hosté	Handicap	Kurs na Chelsea	Kurs na Everton
Chelsea	Everton	0 : 0	1.85	2.05

3.6.5 Systém 1 - 6 nula¹⁵

Jak často se nám stane, že najdeme zápas jasněho favorita s naprostým outsiderem? No, často. Řekl bych, že každý týden jich najdeme několik. Jenže zkušený sázkař se tady do sázení příliš nehrne, v takovýchto "jasných" zápasech bývají kurzy podhodnoceny a ve sportu se přitom, jak víme, může stát cokoli.

Tohle je obecně známo. Vždycky mám pocit marnosti, když vidím, že hráči riskují pro tyto mizerné kurzy svůj cenný kapitál a přitom netuší, že existuje malý a přitom účinný trik, který nám umožní získat na tento zápas kurz na jasněho favorita někde mezi 1,8 až 2,0.

3.6.5.1 Postup:

Místo riskantní sázky velkého kapitálu za nízký kurz vsadíme celou sérii sázek na možné výsledné skóre takového zápasu. Víme přece, jak obvykle podobné zápasy končí. Jedná-li se o zápas domácího favorita s hostujícím outsiderem, očekávejme následující výsledky:

1:0, 2:0, 3:0, 4:0, 5:0, 6:0

Kurzy na přesný výsledek zápasu v takovém případě bývají podobné, jako v následující tabulce:

¹³ Viz [21]

¹⁴ Viz [22]

¹⁵ Viz [23]

<i>Výsledek</i>	1 : 0,	2 : 0,	3 : 0,	4 : 0,	5 : 0,	6 : 0
<i>Kurs</i>	8,0	7,0	9,0	12,0	20,0	50,0

Naše sázky by v tom případě měly být rozloženy následujícím způsobem:

<i>Výsledek</i>	1 : 0,	2 : 0,	3 : 0,	4 : 0,	5 : 0,	6 : 0
<i>Kurs</i>	8,0	7,0	9,0	12,0	20,0	50,0
<i>Částka</i>	65	70	55	40	20	10

Tímto jsme dosáhli toho, že nepříznivý kurz 1,1 či 1,2 byl nahrazen kurzem kolem 2.0.

3.6.5.2 Možné finanční efekty:

Náklady:

V našem případě $65 + 70 + 55 + 40 + 20 + 10 = 260$ Kč

Výnosy:

Pokud získáme kurz kolem dvou, pak tedy bude výnos činit kolem 500 Kč.

(Samozřejmě musíme odpočíst náklady - max. 260 Kč)

3.6.6 Sázení na hodnotu, valuebetting¹⁶

Jedna z cest k systematickému v sázení vede přes tzv. valuebetting, což v překladu znamená, sázení na hodnotu. Neboli nákup hodnotných kurzů. V tomto odvětví máme možnost vydělávat proto, že nám bookmakeři nabízejí špatně nastavené kurzy. Každý kurz odpovídá určitému procentu pravděpodobnosti. Hodnotný kurz je takový, který vyjadřuje pravděpodobnost výsledku menší, než reálnou. Kdyby byli bookmakeři schopni vždy nastavit kurzy, které odpovídají realitě, měli bychom nulovou šanci na zisk v dlouhodobém trendu. Odhalovat jejich chyby je pro nás jedinou možností. Zisk vniká ze sázek, které nám vyjdou. Když budeme sázet pouze na hodnotné kurzy, pak nám budou vycházet častěji, než by odpovídalo pravděpodobnosti vyplývající z nabízeného kurzu. Uvedu malý příklad. Dlouhodobý trend vyjádřím počtem sázek třeba 1000. Hraje se zápas a podle mě je pravděpodobnost, že vyhraje domácí 60%. To odpovídá kurzu 1.67, který tedy subjektivně považuji za správný. Pokud bookmaker třeba z nedostatečné informovanosti vypíše kurz 2, což znamená, že domácí by měli jen 50% naději na úspěch, na kurz vsadím. Ilustrativně, když 1000x za sebou vsadím přesně tento typ zápasu, stane se tohle:

- $1000 \times \text{vklad } 100,- \times \text{nabízený kurz } 2 \times 0.6(\text{reálné procento případů kdy to vyjde}) = 120000,-$.
Kdybych vsadil v kurzu, který by byl nastaven správně tedy 1.67, vypadalo by to následovně:

¹⁶ Viz [24], [25], [7]

- $1000 \times \text{vklad } 100,- \times \text{ nabízený kurz } 1,67 \times 0,6(\text{reálné procento případů kdy to vyjde}) = 100000,-$.

Z toho jasně vyplývá, že když budu dlouhodobě sázet na kurz, který není hodnotný, ale správně nastavený, někdy vyhraju, někdy prohraju, ale za rok budu na nule, samozřejmě, pokud nebudu dokonce sázet podkurzy.

Naopak pokud budu dlouhodobě sázet na konkrétně takto hodnotné kurzy, skončím nakonec s velmi slušným ziskem 20000,-. V tomto případě to odpovídá 20% yieldu, neboli to vyjadřuje, že jsme dlouhodobě schopni sázet na příležitosti s 20% hodnotou.

Na závěr je nutné dodat, že to jestli je kurz od bookmakera správný nebo ne určujeme subjektivně. Vycházíme z informací, které o zápase máme a následně posuzujeme, zda je rozumné zápas v nabízeném kurzu vsadit nebo ne. Jedná se samozřejmě o dost náročnou disciplínu, pro jejíž úspěšné zvládnutí nemá každý člověk dispozice. To, zda jsou naše názory na kurzy správné, nám za čas prozradí statistika, kterou si vedeme. Pokud jsou naše čísla kladná, je to jakýsi důkaz o naší bystrosti a schopnosti objektivity.

3.6.6.1 Nyní uvedu několik poznámek a doporučení:

- Pokud je to možné, sázejte na příležitost, jejíž pravděpodobnost jste určili alespoň na 50%. Toho dosáhnete pomocí asijských handicapů.
- Sázejte až tam, kde vidíte hodnotu alespoň 5%, ale lépe výše, kolem 10%. Eliminujete tím vaši nepřesnost při odhadu pravděpodobností (nikdo si nemůže být 100% jist svým odhadem). U vyšších kurzů sázejte jen na vyšší hodnoty (úměrné výši kurzu).
- Před uzavřením sázky se snažte přemýšlet, jakým směrem se budou měnit kurzy. Kurzy na "všeobecného" favorita obvykle půjdou s blížícím se startem dolů, tudíž kurzy na "všeobecného" outsidera nahoru. Podle toho se rozhodněte, jestli vsadíte raději dříve, nebo počkáte a vsadíte chvíli před startem utkání.

Každá technika má své místo v portfoliu nabídek. Například valuebetting je vhodný převážně na sázení proti favoritům. V zásadě nejde o chyby bookmakerů (těch není mnoho), jako spíš o strategie, které používají. Snaží se například zmást sázkaře tím, že na favorita dají velmi nízký kurz, lidé to pak budou brát jako téměř 100% šanci výhry, přisadí to k nějakému kombinovanému tiketu (jistě navýšení kurzu) a přijdou o celou sázku.

Za zmínku také stojí, že kurzy nejsou 2:2, ale spíš 1.9:1.9, takže při dlouhodobějším sázení s nulovou bilancí neskončím.

3.6.7 Jak sázet a vyhrávat¹⁷

3.6.7.1 Klasický způsob

Nejčastější sázky padají na jasného favorita zápasu. Nejoblíbenější sázky jsou ty s kurzem 1,2 až 1,3. U nižších kurzů není výhra nijak zvlášť zajímavá ve srovnání s existujícím rizikem, takže příliš oblíbené nejsou.

Prohra favorita s takto nízkým kurzem samozřejmě není příliš pravděpodobná, ale může se tak stát. V praxi je tato skutečnost natolik častá, že pokud budete stále stejnou částku sázet na zápasy s kurzem 1,3, tak se postupně prosazíte do mínusu.

3.6.7.2 Jak sázet na kurzy 2,0

Pokud vás děsí výše nutných sázek v případě získávání zpět dříve prohraných, není marné zkusit štěstí u sázek, kde není favorit tak jednoznačný. Ten má v tomto případě kurz zhruba 2, remíza a výhra outsidera má kurz 3. Pořád zde máme jeden favorizovaný tým a navíc mnohem zajímavější částku k výhře. Pokud Váš tým vyhraje, přijmete si na dvojnásobek vsazené částky (za vsazenou stokorunu dostanete dvě, zisk je tedy 100Kč). Pokud ne, chtělo by to zmíněnou stokorunu vyhrát zpátky. V tom případě vsadíte dvojnásobek dříve prohrané částky, opět s kurzem 2. Při výhře za 200Kč dostanete zpátky 400Kč. Od nich odečtete právě vsazených 200Kč, dříve prohraných 100Kč a jste 100Kč v plusu. Pokud prohrajete i podruhé, vsadíte 400 Kč (tedy opět dvojnásobek naposledy prohrané částky).

3.6.7.3 Jak sázet na kurz 3,0

Pokud sázíte na kurz 3 a vyšší, sázíte na outsidera zápasu. Zdánlivě je to nelogické, ale právě protože i outsideri mohou vyhrát, jsou sázkové kanceláře schopny existovat. Je nutné si uvědomit, že pokud jeden tým má kurz 2 a druhý 3, rozdíl mezi nimi není nijak dramatický. Jednoznačný favorit je v případě kurzu 1.2 a nižším. Pokud však má jeden tým 2 a druhý 3, jsou jejich šance srovnatelné. Avšak případná výhra už je o něčem jiném. Pokud sázíte zmiňovanou stokorunu, zisk v prvním případě je 100 Kč, ale v druhém 200 Kč.

Hlavní výhoda sázení na kurz 3 oproti kurzu 2 je pomalý nárůst sázek, pokud se rozhodnete "opravovat" předchozí neúspěšné sázky. V tomto případě totiž nenásobíte předchozí prohranou sázku dvěma ale pouze 1.5 krát, což v praxi velmi oceníte. Posloupnost totiž není 100, 200, 400, 800, 1600, 3200... ale 100, 150, 225, 337.5, 506.3, 759.4. Z jiného pohledu, pokud prohrajete šestkrát po sobě, nejste v tom celkově už za 6300Kč, ale jen za 2078.2Kč. A sedmá sázka není za 6400Kč, ale za 1139.1 Kč.

¹⁷ Viz [26]

Při kurzu 3 můžete mnohem déle neúspěšně sázet, než Vám sázená částka přijde příliš vysoká, nebo ztratíte trpělivost. A pokud máte nějakou zkušenost se sázením, tak sami víte, že reálně vychází téměř každá druhá sázka.

3.6.8 Recency Effect (Vliv nedávné minulosti)¹⁸

Ve sportovním sázení je vizuální obraz vším. Zápasy přenášené televizí po celé zemi shlédne spousta lidí, a proto jsou utkání doprovázená televizním přenosem obvykle ovlivňována sázející veřejností. Jinými slovy, kdyby tým utrpěl ponižující porážku, lajna (line) takového týmu pro jejich příští zápas bude soft (příznivá k tomu, aby ji pokryli). A naopak, pokud tým dosáhne velkolepého vítězství v národním televizním přenosu, jejich hra bude negativně shaded (bude pro ně těžší pokrýt svůj příští zápas).

Toto je ve větším souladu s psychologií sázečích kanceláří než cokoliv jiného. Oddsmakeři nejenom stanoví lajnu pro zápasy; nejenom že handicapují týmy, na které umisťují svoje lajny, ale také handicapují lidi, kteří na jejich lajny sázejí. Například hra, na kterou se nejvíce sází, je Super bowl. Super bowl line je možná ta nejvíce analyzovaná lajna v celé oblasti sportovního sázení. Tato lajna je odrazem jak zapojeného týmu, tak sázející veřejnosti (ale mnohem víc to druhé). Protože oddsmakeři vědí, že budou muset absorbovat spoustu akcí, musí být lajna slušná, aby na obou jejich koncích došlo k vyvážené akci (ne-li tím více na prohrávající straně).

Blowouts (drtivé prohry) při zápasech v národních televizních přenosech, také přináší veřejnosti smysl pro úžas. Veřejnost miluje sázení na vítěze, a jaký existuje lepší způsob, než stát za týmem, který právě utrpěl blowout. Obecně sportovním sázkařům vyhovuje takový tým fadeovat, zvláště pokud to bylo v národním televizním přenosu. Z psychologického hlediska obvykle platí, že tým, který právě vyhrál, se bude cítit nadřazený a bude se chovat trochu arogantně. Budou tedy hrát svou další hru s nádechem arogance a to obyčejně povede k tomu, že nepokryjí a zápas prohrají.

Týmy, které byly právě rozdraceny, obvykle nasazují v příštím zápase větší výkon, obzvláště pokud byl prohraný zápas v televizním přenosu. Tyto týmy cítí potřebu se změnit a následkem toho pak mají lepší výkon. To ve skutečnosti funguje napříč spektrem u všech sportů, ale nejsilněji u NBA (National Basketball Association).

Recency effect je dalším z úhlů pohledu, které se zabývají minulými výkony, a často se používá k předpovídání budoucích výsledků.

¹⁸ Viz [13]

3.6.9 Potenciál a kapacita v kurzovém sázení¹⁹

Ve sportovním handicappingu musí sázkař chápat koncepty potenciálu a kapacity. Jejich definice ve sportovním sázení zní, že potenciál je motivační aspekt týmu k dobrému výkonu (obvykle k výhře). Kapacita je schopnost takového týmu vyhrát. Většina týmů má potenciál vyhrát mnoho zápasů, ale většina z nich k tomu nemá kapacitu. Následující text nám pomůže definovat to, co způsobuje, že má tým potenciál a kapacitu vyhrát zápas.

Situační sázení je nejlepším způsobem, jak sázet na nějaký sport. Přestože zde existuje rutina, protože týmy hrají spoustu zápasů během jakékoliv dané sezóny, nikdo nemůže doopravdy říct, zda ze sebe tým v jakémkoliv konkrétním zápase vydá trochu více námahy. Sportovní sázkař, který používá situační sázení, bude mít proto výhodu. Tento motivační aspekt může být uvažován jako potenciál týmu k dobrému výkonu. Například pokud tým touží po pomstě vůči svému protivníkovi, vyvine velké úsilí a proti svému soupeři nasadí svou nejlepší hru. Rivaly zápasy (proti tradičním rivalům) také spadají do této kategorie. Takové zápasy vzbuzují v hráčích to nejlepší pro jejich fanoušky.

Kapacita týmu skutečně vyhrát zápas je nezávislá na jeho potenciálu k výhře. Jinými slovy, pokud se tým nedokáže vyrovnat svému protivníkovi, žádné množství potenciálu není dostatečné k tomu, aby protivníka porazili. Ve sportovním sázení si musí sázkař uvědomovat, že tým musí mít jak potenciál, tak kapacitu, aby mohl kontrolovat a vyhrát zápas.

Ve skutečném světě sportovního sázení, hlavně v NBA (národní basketbalová asociace), je teorie potenciálu a kapacity testována v praxi. Koncem sezóny nastane mnoho situací, jako dvojitá – a trojitá - pomsta. Týmy jsou vysoce motivovány, aby porazili svého soupeře, ale většinou to nedokáží, protože jednoduše nemohou uplatnit svou pomstu, pokud k tomu nemají kapacitu.

Závěrem lze říct, že potenciál týmu je jeho motivace vyhrát zápas. Kapacita týmu je schopnost týmu zápas vyhrát. Spojení těchto dvou myšlenek přináší skvělé sázecí příležitosti sportovním sázkařům.

3.6.10 Handicapping - jak analyzovat hru nebo zápas²⁰

3.6.10.1 Kdo handicapuje?

Handicapovat může ve sportovním sázení kdokoli, kdo chce mít výhodu. Je potřeba velice pečlivé studování sportovních statistik, abyste mohli přijít s inteligentním odhadem. Někteří lidé to umí velice dobře, zatímco jiní se musí hodně snažit, aby to zvládli. Zajímavá věc na handicappingu je, že lidé mohou přijít s rozdílnými názory, přestože použili ty samé výchozí informace. Proto to záleží čistě a pouze na každém jednotlivci, aby se rozhodnul, které statistiky pro něj budou relevantní a důležité.

¹⁹ Viz [15]

²⁰ Viz [16]

Existují profesionální sázkaři, kteří handicapují za použití sofistikovaných prostředků, například počítačových programů, a pak jsou zde obyčejní sázkaři, kteří sledují denní tisk, aby prozkoumali bodové tabulky z předchozího dne.

Sportovní sázkaři handicapují, aby analyzovali sportovní události a nasměrovali své sázky. Na rozdíl od jiných hazardních her, sportovní sázení doopravdy funguje, pokud je spojeno s dobrou správou peněz a troškou štěstí. Handicapováním člověk zvyšuje svou šanci vyhrát. I když je zde důležitá i pravděpodobnost, která dává konečnou výhodu sázecím kancelářím (v tomto případě těm zabývajícím se kurzovými sázkami), hráči přesto mohou sázkovou kancelář porazit z krátkodobého hlediska. Podobně jako při počítání karet ve hře blackjack, handicapováním člověk vybírá hry, které mají větší pravděpodobnost vyhrát, a to dává sázkaři větší naději, že mu bude pravděpodobnost nakloněna, než kdyby odhadoval výsledek házením mincí.

Žijeme v době internetu a na webu se nachází mnoho sportovních informací. Vynikající zdroj sportovních statistik můžete najít na <http://www.covers.com> (EN). Tyto stránky nabízí sportovním sázkařům rozmanitou škálu statistik pro všechny hlavní sportovní odvětví. Vynikající je na těchto stránkách fórum, kde je všeobecná diskuze a kde sportovní sázkaři debatují o nejnovějších trendech, nabízí si zdarma odhady výsledků apod. Je to skvělé místo, kam můžete něco napsat a necítit se tak osamoceni ve vaší snaze porazit sázkové kanceláře. Dalšími skvělými stránkami, které ale nejsou tolik orientované na sázení, jsou <http://sports.yahoo.com>. Tyto stránky nabízejí bodové tabulky a postřehy o nadcházejících sportovních událostech. Také informují o důležitých zraněních/přestupech sportovců do jiných týmů apod. Všechny tyto stránky jsou také sledovány sázkovými kancelářemi, které informace o zraněních a nejnovějších trendech přijímají s velkým povděkem a používají je hlavně v případě sázek na bodový rozdíl. Týmům, které v současnosti hodně vyhrávají, dávají sázkové kanceláře nevýhodné kurzy.

Handicapping, pokud se uplatňuje správně, je skvělým způsobem, jak získat ve sportovním sázení výhodu. Umožňuje sázkaři mít navrch oproti sázkovým kancelářím, protože může sázet na sportovní události, které podle jeho analýzy vypadají příznivě.

3.6.11 Jak tipovat a čeho se vyvarovat při sázení po internetu²¹

Zde uvádíme pár rad čeho se vyvarovat při sázení po internetu.

3.6.11.1 Několik rad jak tipovat úspěšně

Než začnete se sázením, projděte si desatero rad na úspěšný tip. Poučte se z našich dlouholetých zkušeností a vyhněte se začátečnickým chybám.

- Když podáváte tiket, nedávejte na něho příliš mnoho zápasů (max. 4).

²¹ Viz [10]

- Vyvarujte se tipů na hosty (2). Pokud už na ně dáváte, použijte raději variantu (X2). Kolikrát silné týmy venku zaváhají.
- Při tipování tenisových zápasů se snažte zjistit zdravotní stav hráčů.
- Snažte se zápasy pojistit dvojitipem. Lepší menší kurz a vyhrát.
- Netipujte zápasy, o kterých nemáte dostatek informací, resp. podle kurzu.
- Netipujte přátelské zápasy, ve kterých nejde o nic. Mohou se v nich zrodit mnohá překvapení.
- Netipujte zápasy domácích pohárů. Často se stane, že favorité postaví „béčko“.
- Nehrajte pod vlivem alkoholu nebo léků.
- Hrajte pouze odpočatí a soustředění.
- Vsaďte jen tolik, kolik si můžete dovolit prohrát.

3.6.11.2 Nejčastější chyby při tipování

- Nedostatek disciplíny - podlehnoutí emocím.
- Poučení z proher (analyzujte vaše prohry).
- Touha vyhrát ihned zpět prohrané peníze.
- Nedostatečně prozkoumaná nabídka (používejte statistické tabulky).
- Nepromyšlené sázení ve finanční nouzi. Sázkové kanceláře²²

3.6.12 Jak si zvolit online sázkovou kancelář?

Při sázení po internetu je nejdůležitější výběr sázkové kanceláře. V následující tabulce najdete přehled nejoblíbenějších sázkových kanceláří.

Kancelář	www adresa	Min. sázka	Bonus	Čeština	Sázení zdarma
Sporting bet	www.sportingbet.cz	25Kč	-	ANO	NE
Expekt	www.expekt.com	28Kč	750Kč	ANO	NE
InterWetten	interwetten.com	5Kč	1400Kč	ANO	NE
Unibet	www.unibet.com	27Kč	1000Kč	ANO	NE
Betfair	www.betfair.com	0,1€	20€	ANO	ANO
WorldBet	www.worldbet.com	10Kč	10%	ANO	ANO
Bwin	www.bwin.com	15Kč	1000Kč	ANO	NE
Betway	www.betway.com	2Kč	1000Kč	ANO	NE
Bet-at-home	bet-at-home.com	10Kč	777Kč	ANO	ANO

²² Viz [11]

3.6.12.1 Výhody online sázení oproti klasickým kamenným kancelářím

- Maximální množství sázkových příležitostí.
- Možnost sázet LIVE!
- Bez manipulačního poplatku ze sázení.
- Minimální sázka již od 2,- Kč.
- Dostanete bonus za registraci a první vklad.
- Kurzové sázení z pohodlí domova.
- Kurzové sázení 24 hodin denně, 7 dní v týdnu.
- Kurzové sázení přes mobilní telefon.
- Výsledky vám mohou být preposílány na mobil.
- Jednoduché ovládání, vše v ČEŠTINĚ.

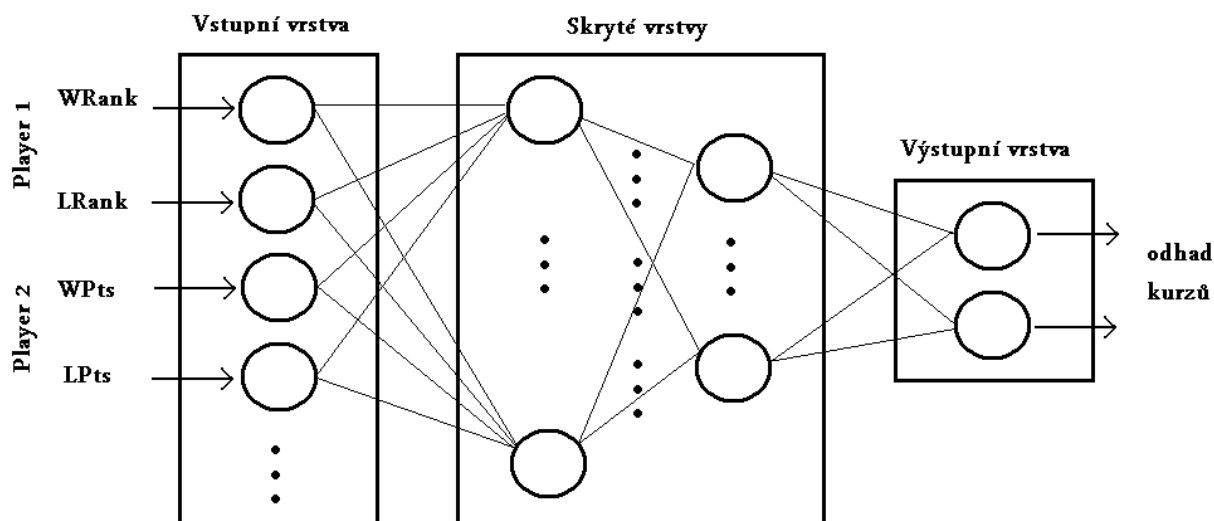
4 Návrh a struktura aplikace

Po prostudování problematiky kurzového sázení a podobných nástrojů pro sázení dostupných na internetu, například OnCourt²³, jsem navrhl nástroj, který pomáhá vybírat vhodné sázkové příležitosti z nabízených kurzů a dalších statických dat. Jako implementační prostředí jsem si vybral NetBeans IDE 6.0 a jazyk Java. Pro testování a hlavně učení aplikace používám výsledky tenisových utkání za posledních několik let.

Aplikace je složena z hlavního balíku agent, ve kterém se nachází třídy pro podporu načítání dat z internetových stránek, práci s databází, konverzí pro konverzi excelovského formátu do MySQL, stahování souborů z internetu, dekompresi stažených souborů, zobrazování dat ve formátu HTML, aj. Část GUI je oddělena od výpočetní. V balíku agent je také knihovna NNlib, které obsahuje třídy implementující neuronovou síť, učící algoritmy jako je backpropagation, a další. Knihovna je navržena i s ohledem na jednoduchost vytváření sítí s různou topologií a je tedy možné nadefinované vrstvy nechat automaticky propojit.

4.1 Neuronová síť

Jako první jsem navrhl neuronovou síť s vícevrstvou topologií a učící se algoritmus backpropagation. Vstupní vektor sítě tvoří 4 až n složek. Jsou to hodnocení hráčů: ranky i body hráčů. Mohou to být také kurzy vyhlášené na vítěze a poraženého několika sázkovými kanceláři, nebo parametry jako je povrch, významnost zápasu, aj. Následuje obrázek topologie sítě, vysvětlení všech zkratk naleznete v části věnující se popisu zdrojových dat. Síť je ve skutečnosti plně propojena, jak je naznačeno.



Obrázek 3 Topologie neuronové sítě

²³ Viz [29]

4.1.1 Učení neuronové sítě

Učení neuronové sítě závisí na několika vlastnostech sítě. První z nich je vhodné propojení, tedy topologie sítě, to se především týká skrytých vrstev. Jestliže jich budete mít příliš mnoho, nebo naopak málo, tak se buď síť vůbec nebude pro dané objemy dat schopná naučit s požadovanou přesností, nebo to bude trvat příliš dlouho. Implementovaná verze s algoritmem backpropagation nijak nemodifikuje koeficient učení alpha, který byl nastaven experimentálně, mezi 0,5 a -0,5. Učení končí, když odezva sítě dosáhne požadované přesnosti ovlivňované chybou sítě. Aby se proces učení nezacyklil, je možné nastavit maximální počet učících cyklů, nebo omezit cyklus učení dobou, po kterou může běžet.

Pro trénování neuronové sítě jsem využil odlehčenou verzi programu upravenou pro předchozí verzi Javy 1.5, kvůli podpoře na výzkumném serveru hele, která dokáže inicializovat síť souborem vzorů, nebo načíst serializovanou, již přetrénovanou síť a dále v jejím tréninku pokračovat. Program umí parsovat parametry z příkazové řádky, které umožňují vytvářet jednoduché skripty, ve kterých se jen parametry určí topologie sítě, koeficient učení a způsob inicializace.

Následuje příklad trénovacího skriptu, který jsem pouštěl na výzkumném serveru **hele.fit.vutbr.cz**.

```
java -jar Agent.jar 0.25 50 50 50 pattern > 0.25_50_50_50.txt &  
java -jar Agent.jar 0.25 100 100 100 pattern > 0.25_100_100_100.txt &  
java -jar Agent.jar 0.25 200 200 200 pattern > 0.25_200_200_200.txt &  
java -jar Agent.jar 0.5 50 50 50 pattern > 0.5_50_50_50.txt &  
java -jar Agent.jar 0.5 100 100 100 pattern > 0.5_100_100_100.txt &  
java -jar Agent.jar 0.5 200 200 200 pattern > 0.5_200_200_200.txt &
```

Jako první se zadává koeficient učení, například 0,5 a pak následuje topologie sítě. Třeba 20 10 5 znamená, že první skrytá vrstva bude mít 20 neuronů, druhá 10 a poslední jen 5. Poslední parametr rozlišuje, zda se síť z inicializuje ze souboru vzorů, nebo ze souboru kam byla síť serializována. Každý z procesů běží odděleně, výsledky jsou přesměrovány do výstupního souboru a po každých n cyklech je síť průběžně ukládána, abychom program mohli kdykoliv ukončit a i tak měli k dispozici nově serializovanou a natrénovanou síť.

Tabulka 1: Výstupem celého procesu mohou být následující soubory

```

/home/xbeloh03/dp/0.25_800.txt - xbeloh03@hele.fit.vutb...
BACKPROPAGATION NET

Creating neuron layers ...OK
Connecting neuron layers ...OK
Reading pattern file ...OK - patterns: 9468

Net structure:
layer 0: 4 neurons
layer 1: 800 neurons
layer 2: 2 neurons
weights: 5602

allright, let's learn...

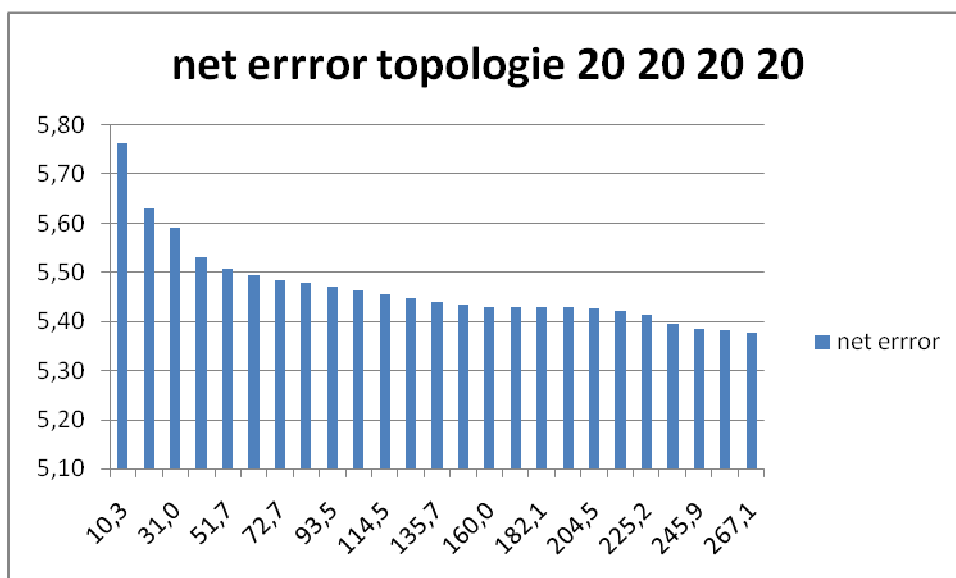
minerror: 0.0050
neterror: 137.65216267800284
accuracy: 2 bad
time    : 386.223 sec
learning rate: 0.25
-----
Řádek: 1/58
    
```

Obrázek 4 textový výstup generovaný při učení neuronové sítě.

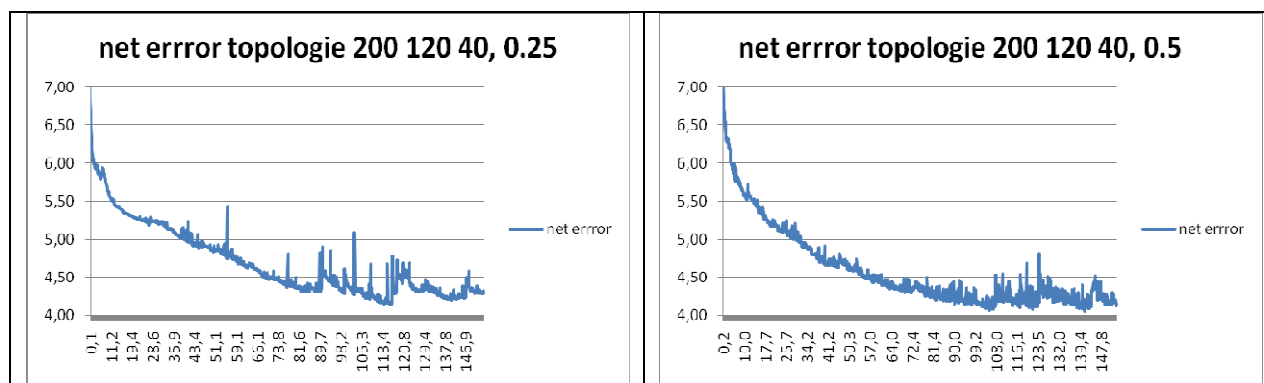
0.25800patternmyobje...	1 203 487
0.25_800.txt	1 076
0.5200200200patternm...	1 502 779
0.5_200_200_200.txt	2 066
Agent.jar	42 160
0.25_300_300_300.txt	293
0.25_1500.txt	442
0.25505050patternmyo...	1 184 779
0.25_50_50_50.txt	24 132
0.5505050patternmyob...	1 184 779
0.5_50_50_50.txt	22 665
0.25100100100pattern...	1 250 779
0.25_100_100_100.txt	6 908
0.5100100100patternm...	1 250 779
0.5_100_100_100.txt	5 887
0.25200200200pattern...	1 502 779
0.25_200_200_200.txt	1 619
0.2520patternmyobject...	1 159 807
0.25_20.txt	20 575
0.520patternmyobject...	1 159 807
0.5_20.txt	20 457
0.25200patternmyobje...	1 169 887

Obrázek 5: znázorňuje výstupní soubory po několika experimentech s topologií sítě a koeficienty učení

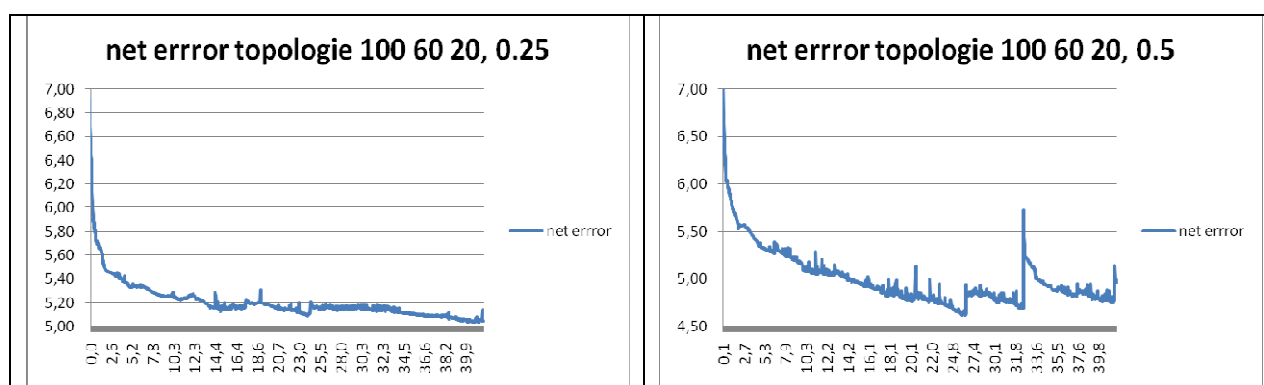
4.1.2 Chyba sítě a přesnost



Obrázek 6 znázorňuje průběh chyby na svislé ose v závislosti na čase v minutách vyneseno na vodorovné ose. Tento průběh je nevyhovující, jelikož přesnost je velmi špatná.



Obrázek 7 znázorňuje podobný průběh závislosti chyby na čase, ale v hodinách. Trénována byla neuronová síť se třemi skrytými vrstvami. V první skryté vrstvě bylo 200, ve druhé 120 a ve třetí 40 neuronů s koeficientem učení 0.25 a 0.5 (s koeficientem 0.75 jsem nedosáhl výrazně lepších výsledků). K trénování jsem využil výzkumného serveru hele.fit.vutbr.cz, kde několik podobných testů (pro různé topologie, počet vrstev, počet neuronů ve vrstvě) bylo spouštěno v průběhu celého týdne.



Obrázek 8: Pro zajímavost uvádím porovnání poloviční topologie, s koeficientem učení 0,25 a 0.5 (s koeficientem 0.75 jsem nedosáhl výrazně lepších výsledků). Čas je uváděn v hodinách na vodorovné ose.

Neosvědčili se ani topologie 30, 30, 30, 30. Teoreticky by ale měla stačit jen jedna skrytá vrstva s dostatečným počtem neuronů. Například pro 1500 neuronů v jediné skryté vrstvě jsem dosáhl hned v začátcích přesnosti sítě kolem 40% a net error pod 0.7

4.2 Zdrojová data²⁴

Jako učící a testovací data slouží zápasy odehrané v předchozích letech. Jsou uloženy v relační databázi MySQL a přistupuji k nim z přes mysql-connector-java-5.1.5.

Zdrojová data z k náhledu tabulky z MS Excelu:

²⁴ Viz [39]

ATP	Location	Tournament	Date	Series	Court	Surface	Round	Best of	Winner	Loser	WRank	LRank	WPts	LPts
1	Adelaide	Next Generation	1.1.2007	International	Outdoor	Hard	Round Robin	3	Luczak P.	Hrbaty D.	166	21	245	1238
1	Adelaide	Next Generation	1.1.2007	International	Outdoor	Hard	Round Robin	3	Serra F.	Johansson J.	60	193	665	202
1	Adelaide	Next Generation	1.1.2007	International	Outdoor	Hard	Round Robin	3	Jones A.	Hajek J.	200	76	194	509
1	Adelaide	Next Generation	1.1.2007	International	Outdoor	Hard	Round Robin	3	Guccione C.	Becker B.	153	58	287	686
1	Adelaide	Next Generation	2.1.2007	International	Outdoor	Hard	Round Robin	3	Mathieu P.H.	Hernych J.	55	75	723	511
1	Adelaide	Next Generation	2.1.2007	International	Outdoor	Hard	Round Robin	3	Djokovic N.	Hajek J.	16	76	1380	509
1	Adelaide	Next Generation	2.1.2007	International	Outdoor	Hard	Round Robin	3	Goldstein P.	Hrbaty D.	70	21	526	1238
1	Adelaide	Next Generation	2.1.2007	International	Outdoor	Hard	Round Robin	3	Gasquet R.	Mayer F.	18	56	1365	700
1	Adelaide	Next Generation	2.1.2007	International	Outdoor	Hard	Round Robin	3	Spadea V.	Stepanek R.	73	19	514	1340

W1	L1	W2	L2	W3	L3	W4	L4	W5	L5	Wsets	Lsets	Comment	B365W	B365L	CBW	CBL	EXW	EXL	PSW	PSL	UBW	UBL
7	6	6	3							2	0	Completed	2,87	1,36	2,85	1,4	2,75	1,42	3,09	1,44	2,77	1,42
2	6	6	3	6	4					2	1	Completed	3	1,33	2,85	1,4	2,53	1,48	2,92	1,47	2,65	1,45
6	3	6	2							2	0	Completed	2	1,72	1,93	1,83	1,95	1,8	2	1,91	2,05	1,73
7	5	6	3							2	0	Completed	2,37	1,53	2,3	1,6	2,3	1,6	2,56	1,6	2,17	1,65
4	6	6	3	6	3					2	1	Completed	1,5	2,5	1,55	2,4	1,5	2,45	1,48	2,87	1,47	2,6
6	0	6	1							2	0	Completed	1,08	6,5	1,11	6,25	1,07	7,19	1,14	7,6	1,09	6,7
6	2	7	6							2	0	Completed	2,87	1,36	3,2	1,33	3,1	1,35	3,51	1,37	3	1,37
6	2	6	1							2	0	Completed	1,19	4	1,22	4,1	1,2	4,25	1,28	4,31	1,22	4,1
6	4	6	4							2	0	Completed	2,62	1,44	2,7	1,45	2,21	1,62	2,67	1,57	2,65	1,45

4.2.1 Zdrojová data popis

Všechna data byla k dispozici v originále v xls formátu.

Legenda k vysvětlení zkratk:

- ATP = Tournament number (men)
- WTA = Tournament number (women)
- Location = Venue of tournament
- Tournament = Name of tournament (including sponsor if relevant)
- Data = Date of match
(prior to 2003 the date shown for all matches played in a single tournament is the start date)
- Series = Name of ATP tennis series
(Grand Slam, Masters, International or International Gold)
- Tier = Tier (tournament ranking) of WTA tennis series.
- Court = Type of court (outdoors or indoors)
- Surface = Type of surface (clay, hard, carpet or grass)
- Round = Round of match
- Best of = Maximum number of sets playable in match
- Winner = Match winner
- Loser = Match loser
- WRank = ATP Entry ranking of the match winner as of the start of the tournament
- LRank = ATP Entry ranking of the match loser as of the start of the tournament
- WPts = ATP Entry points of the match winner as of the start of the tournament
- LPts = ATP Entry points of the match loser as of the start of the tournament
- W1 = Number of games won in 1st set by match winner
- L1 = Number of games won in 1st set by match loser
- W2 = Number of games won in 2nd set by match winner
- L2 = Number of games won in 2nd set by match loser
- W3 = Number of games won in 3rd set by match winner
- L3 = Number of games won in 3rd set by match loser
- W4 = Number of games won in 4th set by match winner
- L4 = Number of games won in 4th set by match loser
- W5 = Number of games won in 5th set by match winner
- L5 = Number of games won in 5th set by match loser
- Wsets = Number of sets won by match winner
- Lsets = Number of sets won by match loser
- Comment = Comment on the match
(Completed, won through retirement of loser, or via Walkover)

Legenda ke kurzům:

- B365W = Bet365 odds of match winner
- B365L = Bet365 odds of match loser
- B&WW = Bet&Win odds of match winner
- B&WL = Bet&Win odds of match loser
- CBW = Centrebet odds of match winner
- CBL = Centrebet odds of match loser
- EXW = Expekt odds of match winner
- EXL = Expekt odds of match loser
- GBW = Gamebookers odds of match winner
- GBL = Gamebookers odds of match loser
- IWW = Interwetten odds of match winner
- IWL = Interwetten odds of match loser
- PSW = Pinnacles Sports odds of match winner
- PSL = Pinnacles Sports odds of match loser
- SBW = Sportingbet odds of match winner
- SBL = Sportingbet odds of match loser
- UBW = Unibet odds of match winner
- UBL = Unibet odds of match loser

4.2.2 Zdrojová data pro neuronovou síť

Jak lze vyčíst z obrázku 3 topologie neuronové sítě, pro její učení a testování je použit jen vektor o několika složkách v porovnání s daty které máme k dispozici. Analýza kompletní tabulky, kterou jsem provedl v rámci předmětu ZZN ukázala, které atributy mají vyšší informační hodnotu vzhledem k predikci kurzů. Jsou to především hodnocení hráčů, jak rank, tak i body které pře turnajem měli. Dále mohou být zahrnuty i odhady dalších kancelářů. Jako další vylepšení se zdá být zajímavé přidat i atributy pokrývající časovou souslednost a vývoj zápasů, typy povrchu, na kterém se hraje a další specifika turnajů (počet setů, významnost turnaje, ...). V neposlední řadě také bilance vzájemných zápasů mezi hráči.

4.3 Databáze

Aplikace si většinu dat nahrává z relační databáze MySQL. Aplikaci jsem testoval s verzí MySQL Windows Essentials (x86) 6.0.3-alpha zpřístupněné přes `mysql-connector-java-5.1.5`. V databázi tennis, která je chráněna heslem a přihlašovacím jménem, si aplikace uchovává seznam hráčů, který je uložen v tabulce `players`. Tabulka `players` má sloupce `Last Name`, `First Name`, `Country`, `Date of Birth` a autoincrementující se `id`. Další tabulkou je tabulka `matches`, kde jsou uchovávány zápasy za poslední roky, měsíce. Tabulka má všechny sloupce, jako mají [zdrojová data](#). Podobně, jako tabulka `players`, má i tabulka `matches` primární klíč autoinkrementující se `id` typu `MEDIUMINT`. Omezení `NOT NULL` vychází již ze samé podstaty klíče typu `PRIMARY`. U ostatních hodnot je povoleno, aby mohly být prázdné.

Není nijak neobvyklé, když v tabulce zápasů, nebo i hráčů chybí některé údaje. Vychází to především z povahy dat. Nikdo přeci na začátku zápasu neví, kolik setů zápas bude mít, nebo jestli bude dokončen. Není ani povinností sázkové kanceláře vypisovat kurzy na všechny zápasy. Někdy jsou chybějící hodnoty zaviněny nedbalostí nebo nekonzistencí předchozích zdrojů. Jak jistě tušíte, důvodů může být hned několik, a tak nemá cenu být příliš restriktivní. Tato problematika souvisí s problémy při čištění a doplňování dat při dolování znalostí, ale tomu byla zasvěcena kapitola o [dolování dat](#).

Pro vytvoření struktury databáze tennis, tedy obou tabulek, slouží následující skript. Pro lepší kompatibilitu a přenosnost jsou všechny sloupce ve formátu `VARCHAR` s délkou 255 znaků, která bývá implicitní, ale pro názornost je ve skriptu vyčíslena. Jedinou výjimkou je pole `Date`, které je typu `DATE`. Používá se pro testování na nejaktuálnější poslední záznam pro upload nových zápasů.

Pro kompletnost je na začátku skriptu vytvořena sama databáze tennis a příkaz `USE` předurčuje, že další operace, jakou jsou vytvoření tabulek, nebo pozdější naplnění budou prováděny právě pro ni.

Záznamy `Winner/Looser` z tabulky `matches` nejsou totožné s tabulkou `players`, proto tu ani není propojení pomocí cizích klíčů. Tabulka `players` slouží jako referenční a přehledová. Data v tabulce `matches` se nedají jednoznačně přiřadit záznamům v tabulce `players`. Je to hlavně způsobeno neúplností záznamů `Winner/Looser` (příjmení a první písmeno z každého jména). Často se stává, že existují hráči se stejným příjmením a podobným jménem.

```
DROP DATABASE tennis;
CREATE DATABASE tennis;
USE tennis;

CREATE TABLE players (
    id MEDIUMINT NOT NULL AUTO_INCREMENT,
    `Last Name` VARCHAR(255), `First Name` VARCHAR(255),
    `Country` VARCHAR(255), `Date of Birth` VARCHAR(255),
    Notes VARCHAR(255), PRIMARY KEY (id)
);

CREATE TABLE matches (
    id MEDIUMINT NOT NULL AUTO_INCREMENT,
    ATP VARCHAR(255), Location VARCHAR(255),
    Tournament VARCHAR(255), `Date` DATE,
    Series VARCHAR(255), Court VARCHAR(255),
    Surface VARCHAR(255), Round VARCHAR(255),
    `Best of` VARCHAR(255), Winner VARCHAR(255),
    Loser VARCHAR(255), WRank VARCHAR(255),
    LRank VARCHAR(255), WPts VARCHAR(255),
    LPts VARCHAR(255), W1 VARCHAR(255),
    L1 VARCHAR(255), W2 VARCHAR(255),
    L2 VARCHAR(255), W3 VARCHAR(255),
    L3 VARCHAR(255), W4 VARCHAR(255),
    L4 VARCHAR(255), W5 VARCHAR(255),
    L5 VARCHAR(255), Wsets VARCHAR(255),
    Lsets VARCHAR(255), Comment VARCHAR(255),
    B365W VARCHAR(255), B365L VARCHAR(255),
    CBW VARCHAR(255), CBL VARCHAR(255),
    EXW VARCHAR(255), EXL VARCHAR(255),
    PSW VARCHAR(255), PSL VARCHAR(255),
    UBW VARCHAR(255), UBL VARCHAR(255),
    LBW VARCHAR(255), LBL VARCHAR(255),
    PSW VARCHAR(255), PSL VARCHAR(255),
```

```

UBW VARCHAR (255) ,UBL VARCHAR (255) ,
CBW VARCHAR (255) ,CBL VARCHAR (255) ,
IWW VARCHAR (255) ,IWL VARCHAR (255) ,
GBW VARCHAR (255) ,GBL VARCHAR (255) ,
SBW VARCHAR (255) ,SBL VARCHAR (255) ,
`B&WW` VARCHAR (255) , `B&WL` VARCHAR (255) ,
PRIMARY KEY (id)
);

```

4.3.1 Práce s databází v programovacím jazyku Java

Následující úryvek kódu je tělem procedury connect(), která se stará o připojení z prostředí Javy do MYSQL. Všimněte si konstrukce vytvoření instance z driveru a vlastního získání spojení přes driver manager. Pro ustavení spojení potřebujeme znát cestu k databázi, přihlašovací jméno a heslo.

```

Class.forName("com.mysql.jdbc.Driver").newInstance();
// připojení k databázi mysql: "jdbc:mysql://localhost/tennis?", "root", "koha8tyx"
conn = DriverManager.getConnection(database_path, database_login, database_password);

```

4.4 Pomocné soubory

Aplikace využívá několik pomocných souborů v různých formátech.

- **settings.txt** je soubor, který obsahuje nastavení programu. Například heslo k databázi, zdroje odkud se stahují soubory z internetu, nastavení parametrů neuronové sítě aj.
- **URL_input.html** slouží k předgenerování zápasů, které se pak zobrazí v okně browseru.
- **skript.sql** je soubor, který obsahuje příkazy SQL, které vytvoří a inicializují databázi tennis MYSQL.
- **PlayerDB.xls** a **2005-8.xls** jsou sešity programu MS Excel stažené z internetu, naleznete tu i původní komprimované verze ve formátu zip, které obsahují databázi hráčů a zápasů.
- **myobject.data** je serializovaná síť, kterou je možno načíst a tím mít k dispozici naučenou síť.
- **input.txt** je tzv. soubor vzorů, který se skládá ze vstupů a kontrolních výstupů, pro které se trénuje neuronová síť. Dalo by se říci, že je vyexportovaným pohledem nad databází zápasů.

4.5 Použité nástroje

Následuje přehled některých nástrojů, programů a vývojových prostředí, které jsem využil:

- Aqua Data Studio 6.5.9, jsem použil pro automatické generování ER diagramů z dostupné databáze, <http://www.aquafold.com> (květen).
- Sada MySQL nástrojů, jak pro skriptovou, tak GUI podporu, pro vytvoření, opravy a další práci s databází: MySQL Server 6.0, Administrator, Migration Toolkit, Query Browser, System Tray Monitor <http://dev.mysql.com/> (květen).
- Navicat 8 je výborný GUI nástroj pro importování databází z MS Excelu a dalších formátů, <http://www.navicat.com/> (květen).
- On Court je program s databází zápasů a různými statistickými informacemi <http://www.oncourt.info/> (květen).
- SAS Enterprise Miner, je podpůrný software pro dolování dat z databází, na kterém jsem ve školních laboratořích v rámci předmětu, získávání znalostí z databází analyzoval databáze zápasů stažených z internetu ve formátu MS Excel <http://www.sas.com/technologies/analytics/datamining/miner/> (květen).
- NetBeans IDE 6.0, je vývojové prostředí. <http://www.netbeans.org/> (květen).

4.6 Další zdroje dat

V předchozích částech práce jsem popsal strukturu perzistentních dat. Aplikace nemá jen možnost tato offline data průběžně aktualizovat, ale také vytvářet si aktuální databáze chystaných zápasů či žebříčků hráčů, pomocí parsování různých informačních portálů na internetu.

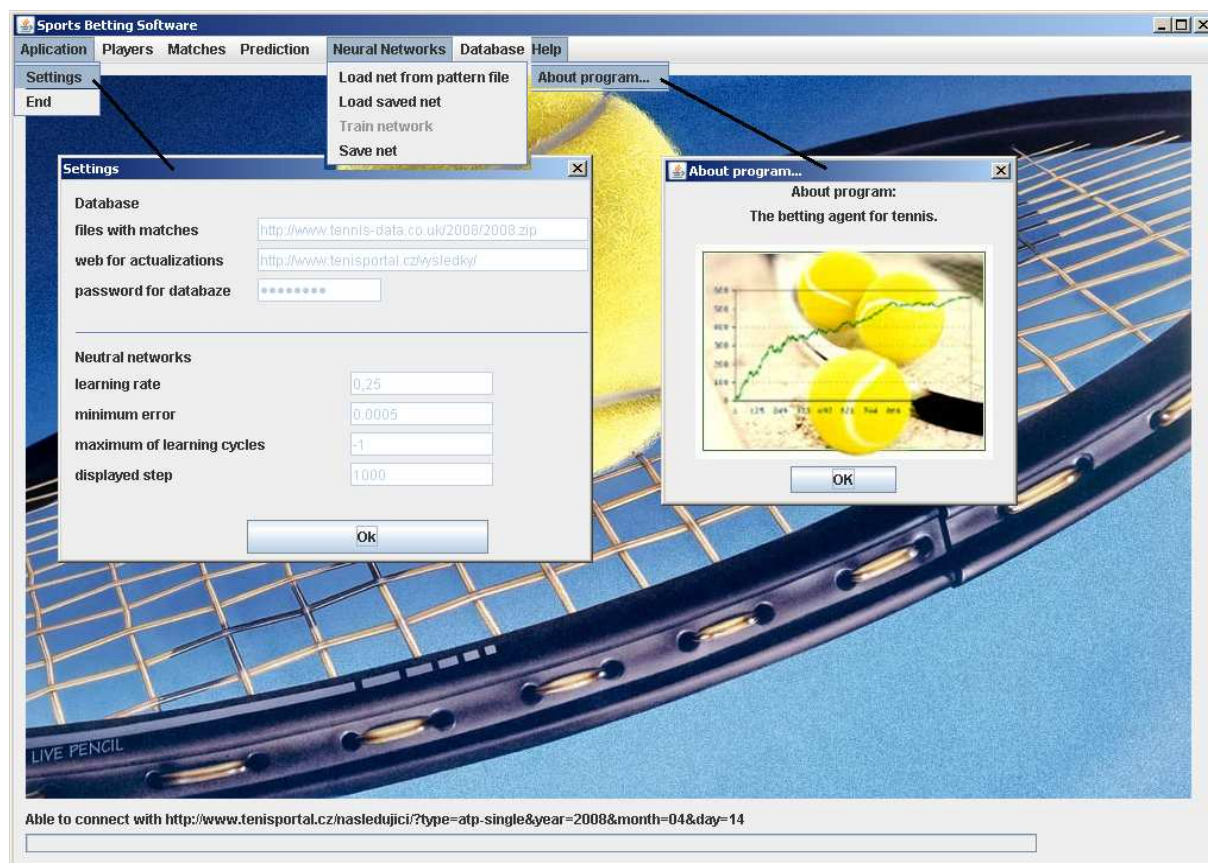
Při načítání stránek z internetu jsem se setkal s několika problémy. Prvním problémem je to, že si nikdy nemůžete být jistí, kdy nám změní strukturu stránky, nebo stránku jen přesměrují či úplně zruší. Časem tedy může aplikace úplně „oslepnout“. Odolnost proti změnám ve struktuře stránky se dá zvýšit použitím schémat a metadat stránky, která nás i při změně odkáží na správné části dokumentu.

Aktuálně aplikace pracuje tak, že vychází z celkem pevně dané struktury stránky. Vše je implementováno ve třídě parsing.java. Která využívá HTML parser k načítání dat z těchto stránek:

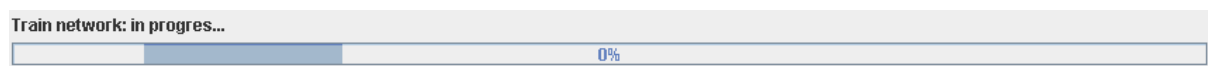
- <http://steveghelper.com/RankingHistory.php>, engine pro vyhledání bodů a ranků hráčů
- <http://sports.espn.go.com/sports/tennis/rankings>, aktuální žebříček hráčů
- <http://www.tenisportal.cz/nasledujici>, odehrané, hrané i plánované zápasy

5 Výsledky

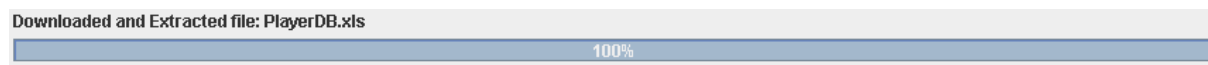
5.1 Implementace v NetBeans IDE 6.0



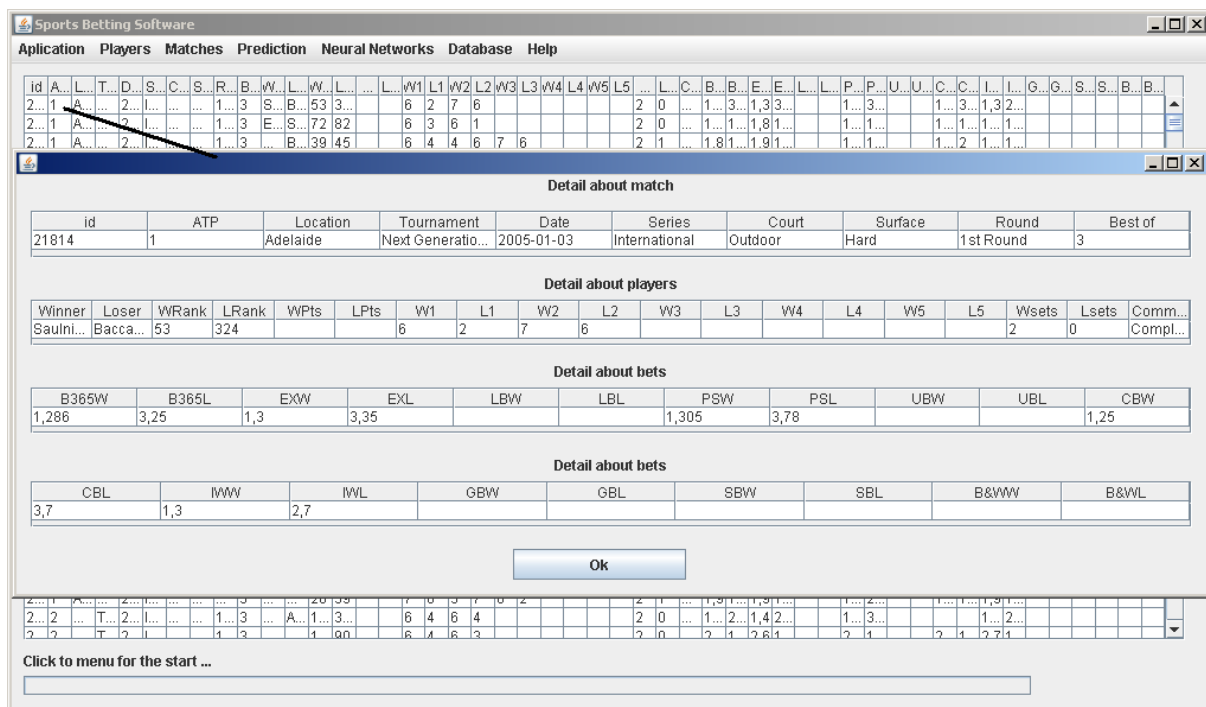
Obrázek 9 ukazuje aplikaci po startu. Uživatel je přivítán tenisovou tematikou na pozadí a dole ve status baru je mu zobrazena nápověda co má udělat. Může například využít bohatého menu aplikace, které mimo níže popsané funkcionality zobrazuje i dialogová okna umožňující nastavení aplikace/databáze a informaci o programu. Například v dialogovém okně Settings můžete nastavit přístupové heslo k databázi v MYSQL. Tato nastavení se načítají a ukládají do speciálního souboru, aby si je program po spuštění „pamatoval“.



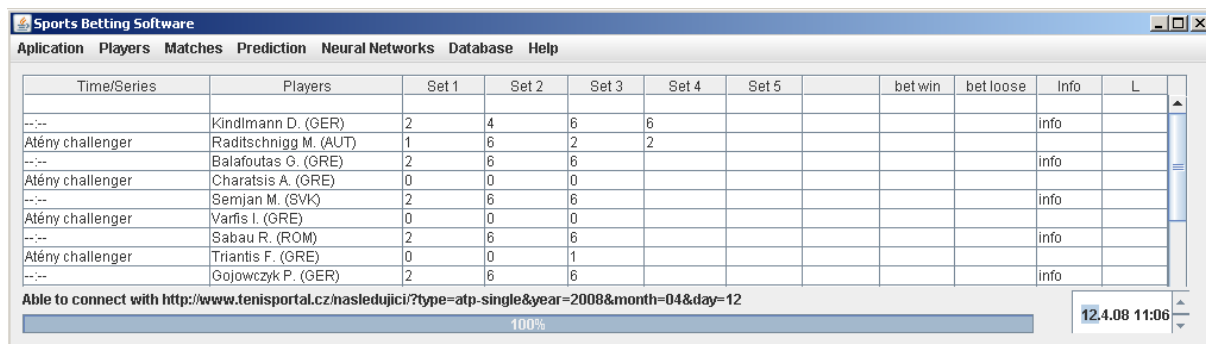
Obrázek 10: Po načtení neuronové sítě ze souboru vzorů nebo při načtení již uložené, předtrénované sítě, má uživatel možnost síť ještě lépe natrénovat. V progress baru vidíte, jak je znázorňován trénink neuronové sítě. Jelikož se jedná o dosti výpočetně náročnou činnost, je samozřejmé, že tento proces běží v samostatném vlákně, aby neblokoval další činnost při práci s aplikací prostřednictvím GUI.



Obrázek 11 Většina datových zdrojů, které se často mění, si aplikace stahuje přímo z internetu, a kopie si nechává pro případ offline režimu ve svém pracovním adresáři pro případ nedostupnosti internetového připojení, nebo výpadku serveru poskytující data. Ve status baru vidíme ohlášení úspěšného stažení databáze hráčů.



Obrázek 12 Po zobrazení zápasů z tabulky matches, vidíme všechny informace uspořádané do sloupců. Pro přehlednější zobrazení všech hodnot, stačí kliknout na příslušný řádek a v dialogovém okně uvidíme přehledně rozdělená data do čtyř tabulek.



Obrázek 13 Tabulka, kterou vidíte, vznikla načtením zápasů pro den, který se dá zvolit v pravém dolním rohu. Implicitně je načten aktuální den. Vidíme zápasy, které se budou dnes hrát, a můžeme si na ně případně vsadit.

Odhad kurzu

výber z odehraných zápasů

Querrey S. x Darcis S. Odhad

Player1	Player2	WRank	LRank	WPts	LPts	B365W	B365L	CBW	CBL	EXW	EXL	PSW	PSL	UBW	UBL	LBW	LBL
Querrey...	Darcis S.	63.0	86.0	573.0	474.0	1.53	2.37	0.0	0.0	1.63	2.2	1.667	2.4	1.6	2.3		

výber z plánovaných zápasů

Poch-Gradin C. (ESP) x Traykov I. (BUL) Odhad

Match	Winner	Looser	w bet	l bet
Atény challenger	Poch-Gradin C. (ESP)	Traykov I. (BUL)	null	null

zadej vlastní zápas Odhad

Player 1 surname	Player 1 first name	Player 2 surname	Player 2 first name
Stepanek	Radek	Del Potro	Juan Martin

odhadovavy kurz na zápas

Player 1 Player 2 Ok

Obrázek 14 zobrazuje okno, ve kterém je možné provádět odhady kurzů. Vidíme tu tři tlačítka pro odhad. Horní slouží k odhadu kurzů pro zápasy z databáze zápasů, které se použila pro učení sítě a je ji možné průběžně aktualizovat. Prostřední slouží k odhadu kurzů zápasů, které se teprve hrát mají a dolní k odhadu na vlastní zápas. Výchozí informací jsou jména hráčů, podle kterých se přes internetové vyhledávací servery zjišťují aktuální žebříčky, body hráčů a další statistiky. Jestliže je jméno hráče nepřesné, nebo k němu nejsou dostatečné statistiky, objeví se dialogové okno, ve kterém je možné chybějící hodnoty doplnit. Viz následující obrázek.

Poch-Gradin C. (ESP)

I can't find rank and points of the player.

Please insert values:

Rank	Points
<input type="text" value="64"/>	<input type="text" value="480"/>

Insert

Poch-Gradin C. (ESP)

I can't find rank and points of the player.

Please insert numbers!

Rank	Points
<input type="text" value="newim"/>	<input type="text"/>

Insert

Poch-Gradin C. (ESP)

I can't find rank and points of the player.

Pts has to be in the interval <1...20 000>

Rank	Points
<input type="text" value="13"/>	<input type="text" value="11111111"/>

Insert

Obrázek 15 Dialogové okno si žádá doplnění ranku a bodů pro jméno hráče zobrazené v jeho titulkovém pruhu. Umí samozřejmě ohlásit špatné zadání vstupů.

Rank	Player	Contry	Points
1	Roger Federer	Switzerland	6600.0
2	Rafael Nadal	Spain	5930.0
3	Novak Djokovic	Serbia & Montenegro	4785.0
4	Nikolay Davydenko	Russia	3175.0
5	David Ferrer	Spain	2570.0
6	Andy Roddick	USA	2260.0
7	David Nalbandian	Argentina	2070.0
8	James Blake	USA	1920.0
9	Tomas Berdych	Czech Republic	1875.0
10	Richard Gasquet	France	1855.0
11	Mikhail Youzhny	Russia	1585.0
12	Paul-Henri Mathieu	France	1495.0
13	Marcos Baghdatis	Cyprus	1480.0
14	Fernando Gonzalez	Chile	1445.0
15	Jo-Wilfried Tsonga	France	1440.0
16	Carlos Moya	Spain	1440.0
17	Juan Monaco	Argentina	1440.0
18	Tommy Robredo	Spain	1400.0
19	Juan Carlos Ferrero	Spain	1390.0
20	Lleyton Hewitt	Australia	1375.0
21	Andy Murray	Great Britain	1305.0
22	Ivo Karlovic	Croatia	1300.0
23	Guillermo Canas	Argentina	1255.0
24	Philipp Kohlschreiber	Germany	1175.0
25	Jarkko Nieminen	Finland	1140.0
26	Igor Andreev	Russia	1105.0
27	Nicolas Pietrangeli	Spain	1105.0
28	Radek Stepanek	Czech Republic	1090.0
29	Ivan Ljubicic	Croatia	1085.0
30	Stanislas Wawrinka	Switzerland	1085.0
31	Fernando Verdasco	Spain	1050.0
32	Dmitry Tursunov	Russia	1010.0
33	Gilles Simon	France	940.0
34	Feliciano Lopez	Spain	935.0
35	Juan Ignacio Chela	Argentina	905.0
36	Tommy Haas	Germany	895.0
37	Janko Tipsarevic	Serbia & Montenegro	886.0
38	Robin Soderling	Sweden	830.0

Able to connect with <http://sports.espn.go.com/sports/tennis/rankings>

Obrázek 16 ukazuje žebříček 100 nejlepších hráčů. Je to praktické hlavně tehdy, když si sestavujete vlastní zápas pro významné turnaje, a můžete si hodnoty bodů a ranků velmi rychle porovnat sami pro nasazené hráče. Žebříček je načítán z internetového serveru, viz status bar.

Database
Download database of players
Download database of matches
Load database of players
Load database of matches
Upload databases of matches from all years
Upload new matches
Delete database of players
Delete database of matches
Create database tennis
Run script

Obrázek 17: Menu Database obsahuje příkazy pro práci s databází.

- Můžeme zde stahovat komprimované excelové soubory tabulek players a matches z internetu volbou Download database of
 - Pomocí příkazu Load database of ... se stažené soubory načtou z excelovského formátu do MYSQL databáze.
 - Upload databases of matches from all years, dokáže naplnit databázi výsledky za několik posledních let. Například od roku 2000 až do současnosti.
 - Upload new matches pracuje tak, že vyhledá nejnovější záznam v databázi a načte jen ty zápasy, které jsou novější.
- Delete database of ... smaže obsah vybrané databáze.
 - Create database tennis nebo run skript zajistí provedení skriptu, které byl uveden výše. První možnost spustí příkazy skriptu jednotlivě, jak jsou zapsány v Java kódu, slouží jako defaultní nastavení. Druhá volba umožňuje spouštět skript, který je možné editovat a dělat v něm mírné úpravy.

Atény challenger	S	1	2	3	4	5	K1	K2	
Kindlmann D. (GER)	2	4	6	6					info
Raditschnigg M. (AUT)	1	6	2	2					info
Balafoutas G. (GRE)	2	6	6						info
Charatsis A. (GRE)	0	0	0						info
Semjan M. (SVK)	2	6	6						info
Varfis I. (GRE)	0	0	0						info
Sabau R. (ROM)	2	6	6						info
Triantis F. (GRE)	0	0	1						info
Gojowczyk P. (GER)	2	6	6						info
Triantis V. (GRE)	0	0	2						info
Brown D. (JAM)	2	6	6						info
Champeris S. (GRE)	0	0	3						info
Kapogiannis C. (GRE)	2	6	7						info
Bisbikos P. (GRE)	0	0	5						info
Rovas N. (GRE)	2	6	6						info
Drymaltis G. (GRE)	0	1	0						info
Traykov I. (BUL)	2	6	6						info
Mouratoglou A. (GRE)	0	1	1						info
Luncanu P. (ROM)	2	6	6						info
Tarassis G. (GRE)	0	2	0						info
Van Gernerden M. (NED)	2	6	6						info
Serkedakis N. (GRE)	0	3	2						info
Ianni S. (ITA)	2	6	6						info

Obrázek 18 Tabulky dnešních zápasů lze zobrazit i ve formátu HTML.

```

BACKPROPAGATION NET

Creating neuron layers ...OK
Connecting neuron layers ...OK
Reading pattern file ...OK - patterns: 844

Net structure:
layer 0: 4 neurons
layer 1: 20 neurons
layer 2: 20 neurons
layer 3: 20 neurons
layer 4: 20 neurons
layer 5: 2 neurons
weights: 1402

allright, let's learn...

cycle: 100

input:          target:        output:         error:
63 86 573 474   1.53 2.37      1.54 2.33      0.0046496983
195 30 218 930  6.5 1.1        5.76 1.11      0.07494599
82 44 481 690   3.75 1.25      3.48 1.32      0.034089938
84 150 476 286  1.4 2.75       1.44 2.67      0.0118166935
88 55 468 611   1.72 2.0       1.79 1.97      0.009996837
93 29 460 955   2.5 1.5        2.57 1.49      0.007341814
25 68 1320 540  1.44 2.62      1.45 2.64      0.0039737397
80 131 487 338  0.0 0.0        0.24 0.40      0.06468476
91 126 462 357  1.57 2.25      1.58 2.22      0.004086405
27 224 995 184  1.16 4.5       1.18 4.20      0.0319925
43 76 714 517   1.28 3.5       1.33 3.28      0.027509551
77 53 515 621   2.87 1.38      2.73 1.46      0.022612076

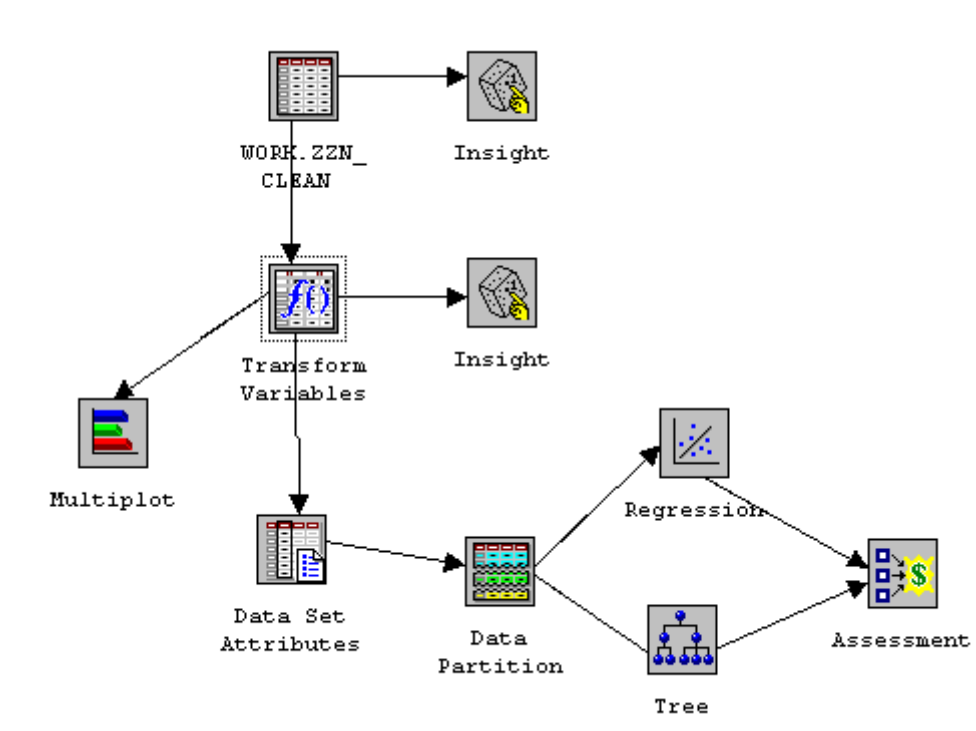
recalling '1 45 6550 753 ': 1.04 8.13
recalling '41 11 781 1545 ': 2.77 1.48
recalling '2 16 5705 1495 ': 1.33 3.36
recalling '4 62 2710 565 ': 1.28 3.88
recalling '6 72 2160 533 ': 1.32 3.44
recalling '2 9 5705 1800 ': 1.38 2.93
recalling '10 32 1685 985 ': 1.51 2.47
recalling '4 41 2710 781 ': 1.34 3.22
recalling '6 1 2160 6550 ': 3.29 1.31
recalling '2 10 5705 1685 ': 1.36 3.06
recalling '4 6 2710 2160 ': 1.68 2.28
recalling '4 2 2710 5705 ': 2.81 1.54

minerror: 5.0E-6
neterror: 0.432821896516907
accuracy: 56.71781034830929%
time : 11.098 sec

```

Obrázek 19 znázorňuje průběh tréninku neuronové sítě. Nejdříve je vidět její topologie. Můžeme tedy vyčíst, kolik skrytých vrstev sítě má. Aby se nezobrazoval každý cyklus, můžeme výpisy omezit, viz nastavení. Nejdříve probíhá učení pomocí algoritmu back propagation a pak v části recalling už jen sít' zkusíme bez zpětné vazby. Nakonec vidíme minimální chyby v síti, chybu sítě, přesnost a také celkový čas učení.

5.2 SAS Enterprise Miner²⁵



Obrázek 20: Diagram

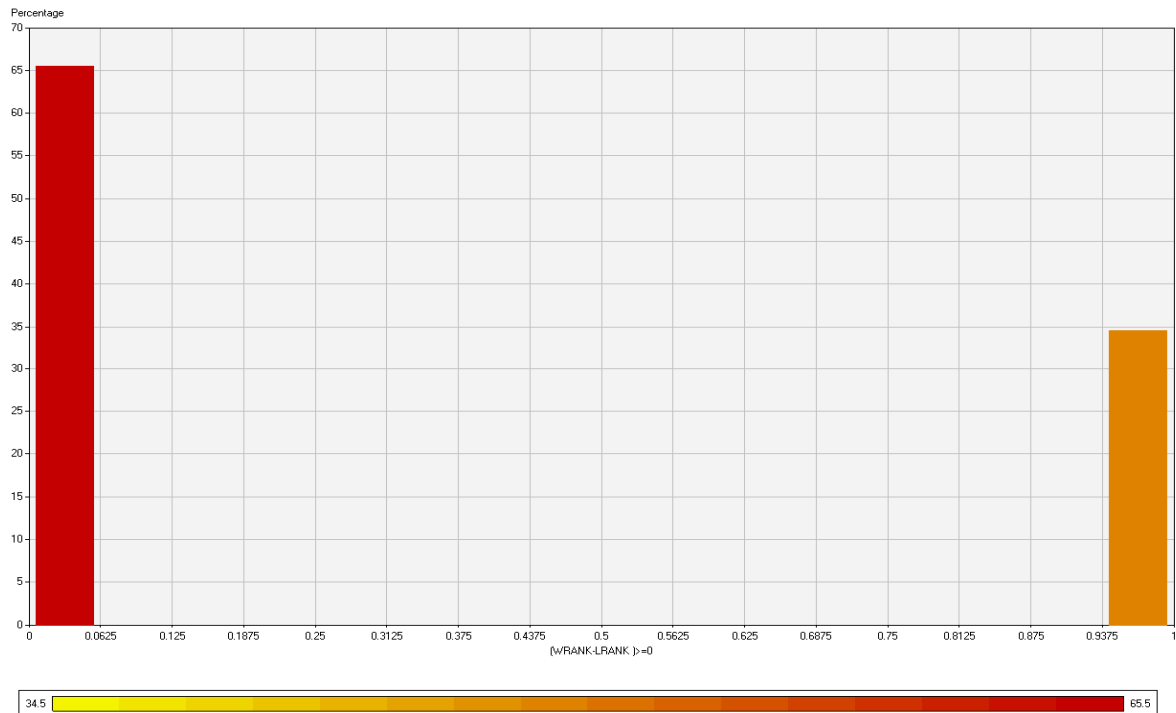
5.2.1 Charakteristika proměnných

Name	Keep	Role	Formula	Mean	Std Dev	Skew	Kurtosis	C.V.	Format	Label
UBL_W	Yes	input	MEAN(UBL-UBW)>=0	0.6405	0.4799738925	-0.586034544	-1.658222737	0.7493737588	BEST12.	MEAN(UBL-UBW)>=0
PSL_W	Yes	input	MEAN(PSL-PSW)>=0	0.671	0.4699674826	-0.72843836	-1.470849406	0.7003986328	BEST12.	MEAN(PSL-PSW)>=0
EXL_W	Yes	input	MEAN(EXL-EXW)>=0	0.654	0.4758121463	-0.647962249	-1.581727653	0.7275415082	BEST12.	MEAN(EXL-EXW)>=0
CBL_CBW	Yes	input	MEAN(CBL-CBW)>=0	0.6475	0.4778681083	-0.61794309	-1.619767106	0.7380202445	BEST12.	MEAN(CBL-CBW)>=0
B365W_L	Yes	input	MEAN(B365L-B365W)>=0	0.665	0.4721085079	-0.699691543	-1.511944691	0.7099376058	BEST12.	MEAN(B365W-B365L)
BET_MEAN_L	Yes	input	MEAN(B365L,CBL,EXL,PSL,UBL)	3.3075973778	2.9450390263	5.0011790412	42.295518262	0.8903861897	BEST12.	MEAN(B365L,CBL,EXL,PSL,UBL)
BET_MEAN_W	Yes	input	MEAN(B365W,CBW,EXW,PSW,UBW)	1.8308363374	0.8841938323	2.6332200687	10.094746285	0.4829453153	BEST12.	MEAN(B365W,CBW,EXW,PSW,UBW)
RANG_DIFF_SIGN	Yes	input	(WRANK-LRANK)>=0	0.345	0.4754871649	0.6526155345	-1.575669635	1.3782236665	BEST12.	(WRANK-LRANK)>=0
PTS_DIFF_SIGN	Yes	input	(WPTS-LPTS)>=0	0.6505	0.4769313176	-0.631749877	-1.60249559	0.7331765067	BEST12.	(WPTS-LPTS)>=0
PTS_DIFF	Yes	input	WPTS-LPTS	443.36715218	1289.899826	2.2720790658	8.698776574	2.9093265472	BEST12.	WPTS-LPTS
RANG_DIFF	Yes	input	WRANK-LRANK	-36.33584337	151.04272486	-1.26400187	24.592307916	-4.156852046	BEST12.	WRANK-LRANK
ATP	Yes	input		34.1505	18.640214015	0.0304078072	-1.157273791	0.5458255081	BEST12.	ATP
WRANK	Yes	input		67.363181591	93.426922002	5.962741446	59.522999996	1.3869137383	BEST12.	WRank
LRANK	Yes	input		103.58153537	129.47540703	4.6524411838	33.240840294	1.2499854009	BEST12.	LRank
WPTS	Yes	input		1155.7133567	1242.3006203	3.2020169143	11.970758511	1.0749210547	BEST12.	WPts
LPTS	Yes	input		713.85040161	596.14023941	2.7920791971	13.678813116	0.8351052799	BEST12.	LPts
B365W	Yes	input		1.7966949153	0.8583130002	2.4201770412	7.9148780064	0.477717721	BEST12.	B365W
B365L	Yes	input		3.1259586864	2.2905189791	3.3992449504	23.53283974	0.7327412832	BEST12.	B365L
CBW	Yes	input		1.8184748513	0.8627171598	2.7097241396	11.022999669	0.4744179768	BEST12.	CBW
CBL	Yes	input		3.1338507301	2.2323005603	2.5101669787	7.6260560753	0.7123187262	BEST12.	CBL
EXW	Yes	input		1.7999784599	0.8257956747	2.4296662127	8.0221777222	0.4587808649	BEST12.	EXW
EXL	Yes	input		3.0340199246	2.0444218639	2.9030027405	14.921572209	0.6738327087	BEST12.	EXL
PSW	Yes	input		1.9240980898	1.0831311607	3.7900576631	24.571482877	0.5629292843	BEST12.	PSW
PSL	Yes	input		3.9111868869	5.0403385857	7.230742176	77.91576113	1.288697966	BEST12.	PSL
UBW	Yes	input		1.7999945534	0.8255832056	2.5909573058	10.152118375	0.4586587243	BEST12.	UBW
UBL	Yes	input		3.2198965142	3.1777809405	6.9209886902	74.726964562	0.9869202089	BEST12.	UBL

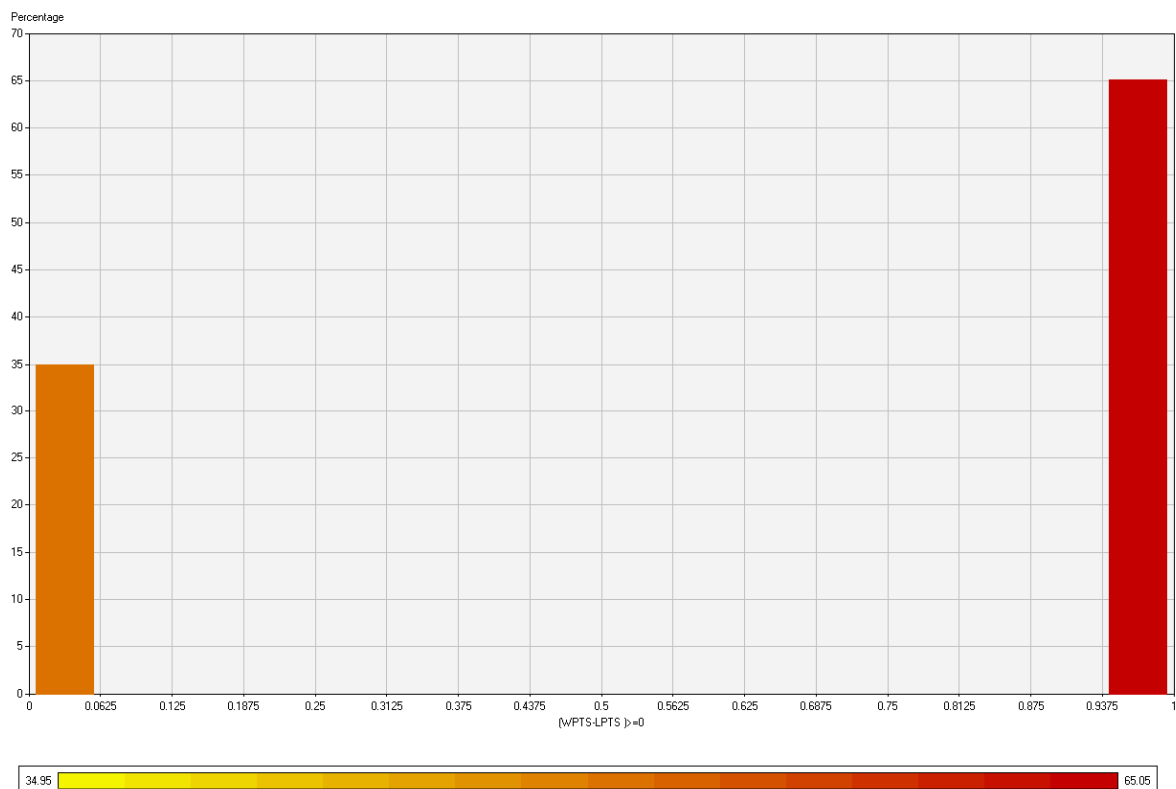
Obrázek 21: Proměnné z uzlu Transform Variables

²⁵ Viz [40]

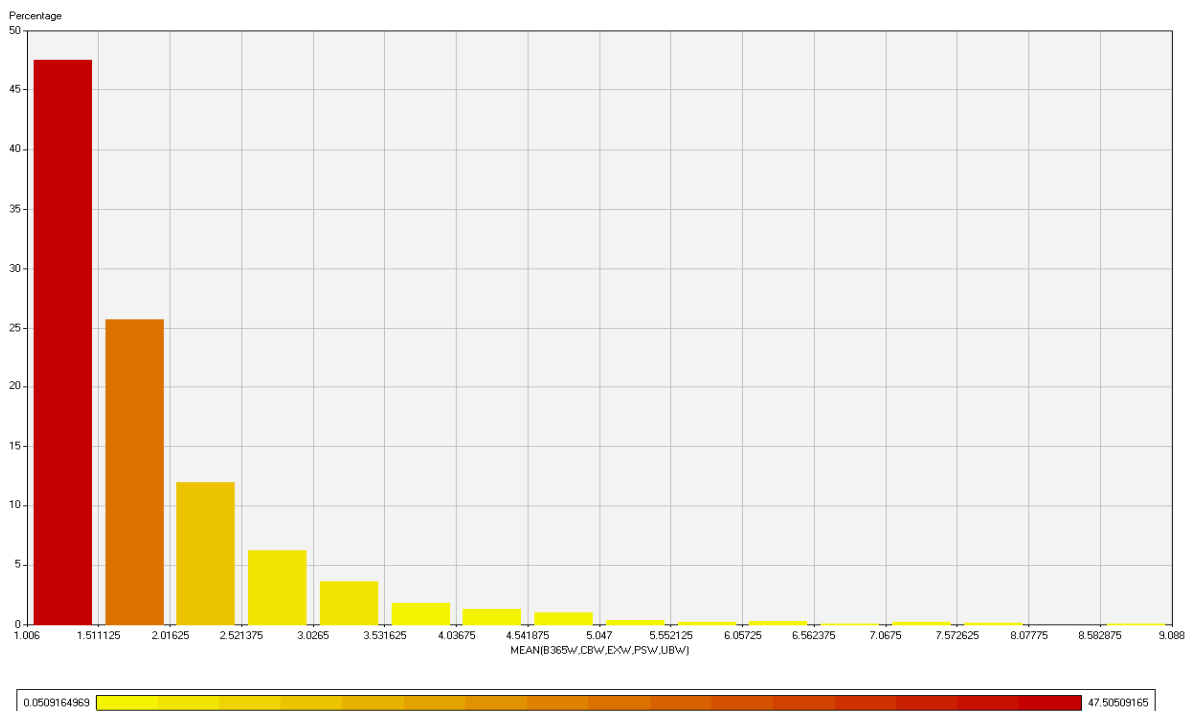
5.2.2 Grafy a histogramy



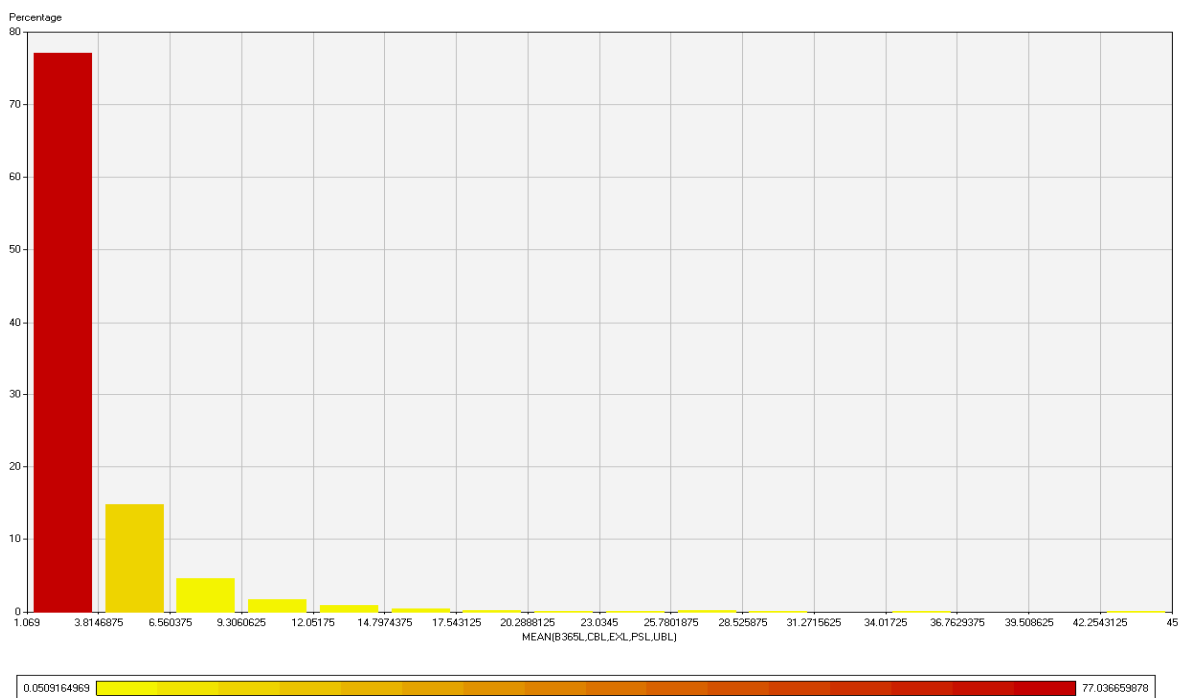
Obrázek 22: $(WRANK-LRANK) \geq 0$
Distribuce proměnné rozdílu ranků nám prozrazuje, že hráči s větším rankem vyhráli v 65,5% případů.



Obrázek 23: $(WPTS-LPTS) \geq 0$
Distribuce proměnné rozdílu bodů nám prozrazuje, že hráči s více body vyhráli v 65,05% případů. V porovnání s předchozím grafem z rozdílu ranků, je vidět že povaha těchto hodnocení je opačná.



Obrázek 24: MEAN(B365W,CBW,EXW,PSW,UBW) Z distribuce průměrů kurzů všech sázkových kanceláří je vidět, že kurzy se z více než poloviny pohybují v intervalu od 1 do 1,51. Významnou skupinu tvoří i kurzy do 2,5. Ostatní jsou již málo časté. To proto, že průměr je z kurzů na vítěze a ideální odhad by se měl blížit 1.

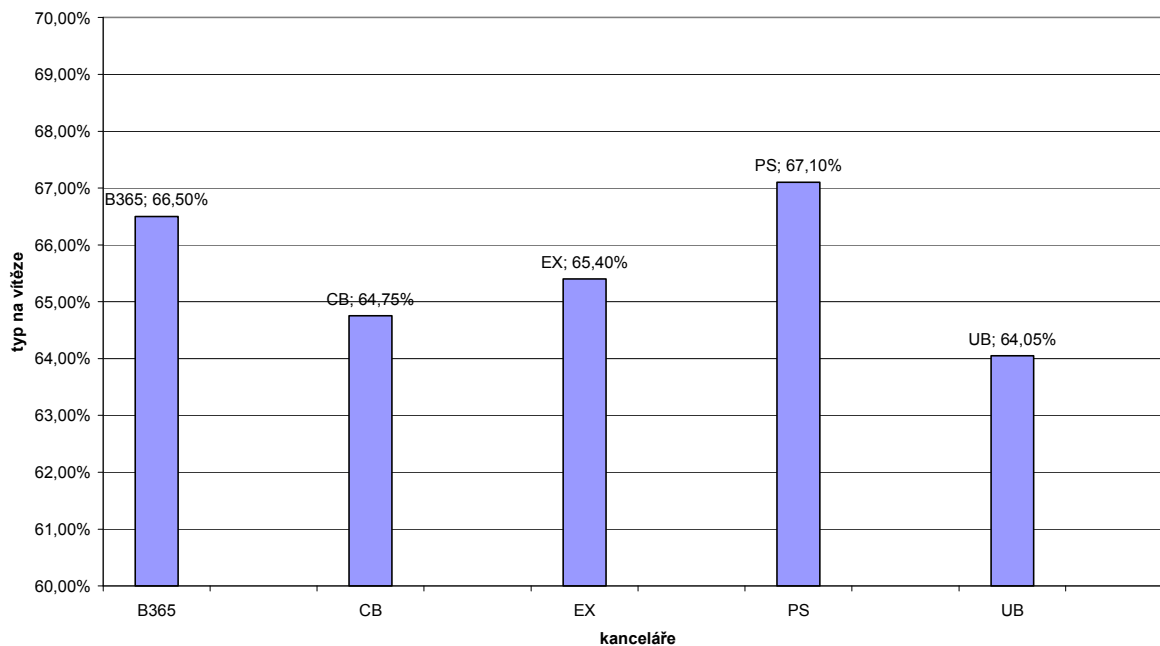


Obrázek 25: MEAN(B365L,CBL,EXL,PSL,UBL) Z distribuce průměrů kurzů všech sázkových kanceláří je vidět, že kurzy se z více než ¼ pohybují v intervalu od 1 do 3,8. Významnou skupinu tvoří i kurzy do 6,5 a 9,3. Ostatní jsou již málo časté. To proto, že průměr je z kurzů na poražené a ideální odhad by se měl blížit ∞ . V porovnání s předchozím grafem vítězů, je z tohoto grafu vidět, že v případě odhadů kurzů poražených se dostáváme na hodnoty kurzů již v prvním sloupci, ale na kurzy okolo 3,8 jsme si museli v grafu vítězů počkat až na šestý sloupec. Z toho plyne, že hodnoty kurzů mají silnou vypovídající hodnotu a kanceláře jsou schopny pomocí kurzů předpovídat vítěze a poražené.

Tabulka 2: Průměry kurzů na vítěze a poražené jednotlivých sázkových kanceláří.

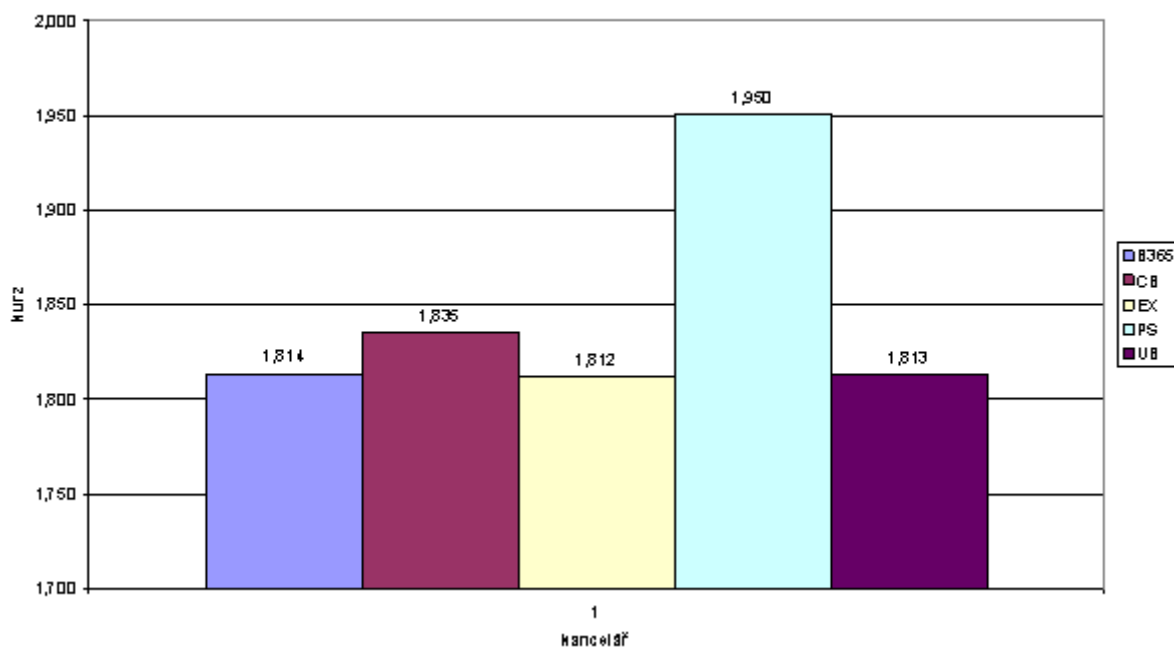
B365		CB		EX		PS		UB	
W	L	W	L	W	L	W	L	W	L
1,813	3,265	1,836	3,302	1,811	3,201	1,950	4,269	1,815	3,452
66,35%		65,15%		65,50%		66,45%		64,20%	

Porovnání odhadu sázkových kanceláří



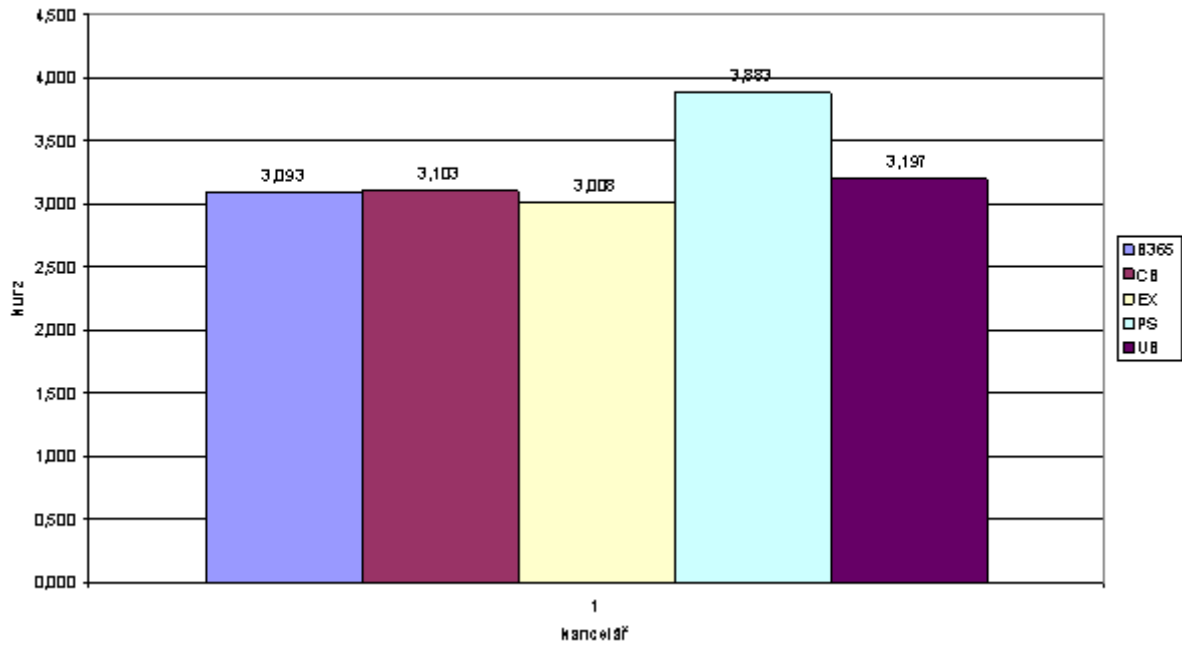
Obrázek 26: Porovnání odhadu kurzů jednotlivých sázkových kanceláří.

Tip na vítěze

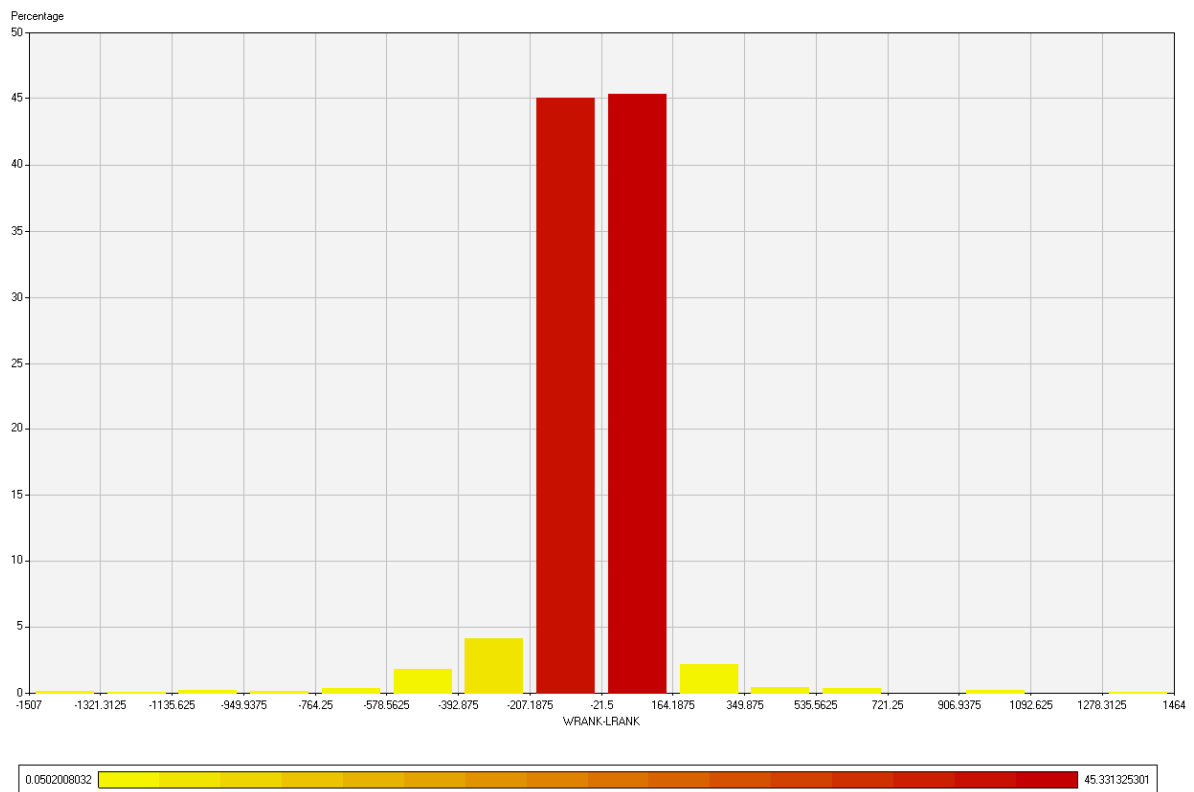


Obrázek 27: Přehled odhadu kanceláří při tipování na vítěze. Jak je vidět, tak se nejméně dařilo kanceláři PS. Ostatní kanceláře mají silně vyrovnané skóre.

Tip na poraženého

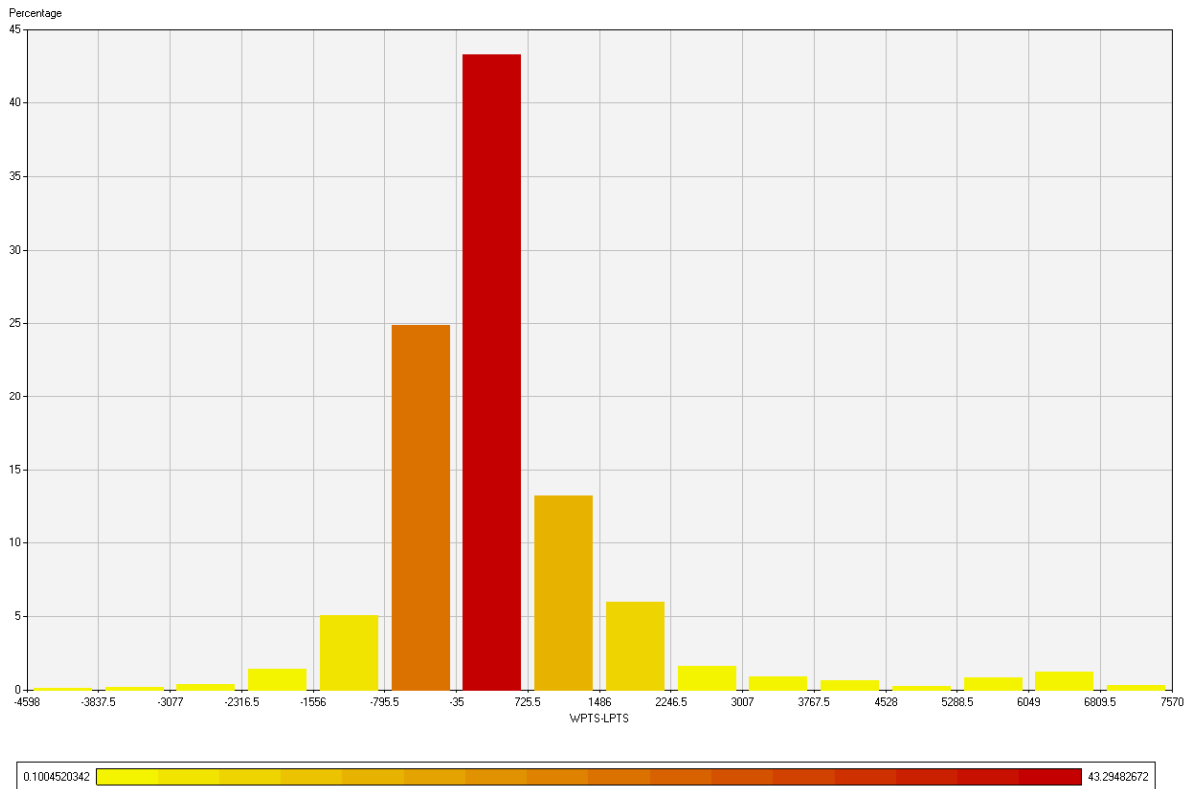


Obrázek 28: Přehled odhadu kanceláří při tipování na poraženého. Jako jedny z nejlepších i pro předchozí odhady jsou B365 a EX, zato PS potvrdila svoji pověst i v tomto grafu.



Obrázek 29: WRANK-LRANK

Distribuce proměnné rozdílu ranku nám prozrazuje, že ve velkých turnajích jsou proti sobě nasazováni hráči nejčastěji s rozdílem ranků v intervalu: -207...164



Obrázek 30: WPTS-LPTS

Distribuce proměnné rozdílu bodů nám prozrazuje, že ve velkých turnajích jsou proti sobě nasazováni hráči nejčastěji s rozdílem bodů v intervalu: -795...725. Graf připomíná nápadně normální rozložení.

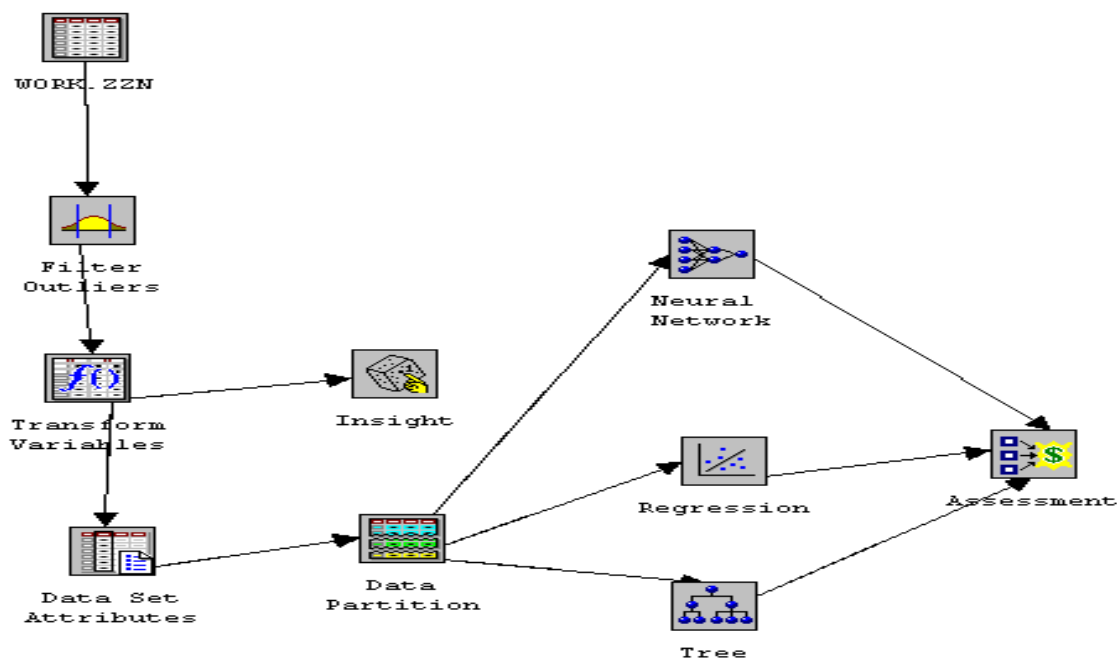
Dolování

Jako vstupní atribut použijeme rozdíl bodů obou hráčů podle vztahu:

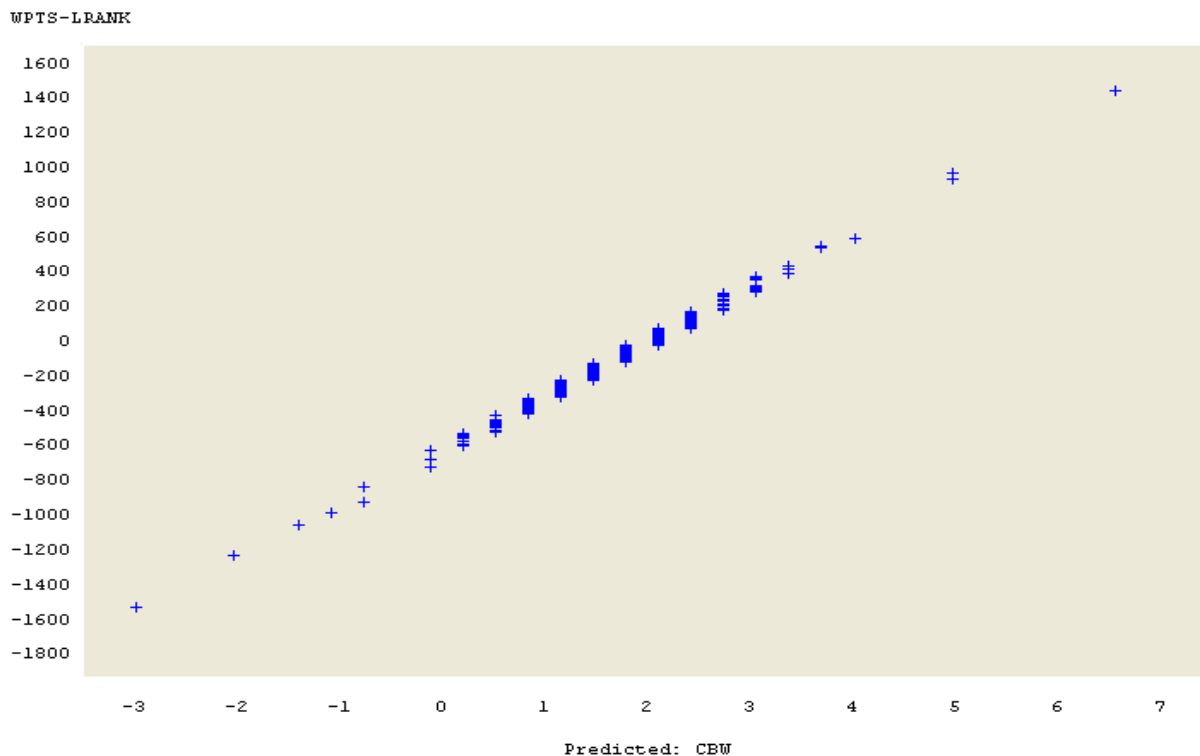
$$WPTS - LPTS.$$

Jako výstupní hodnotu určíme výslednou hodnotu kurzu, kterou stanoví sázková kancelář *CB* na vítěze.

Jako metodu jsem zvolil regresní analýzu



Obrázek 31 – Diagram zapojení komponent v SAS Mineru

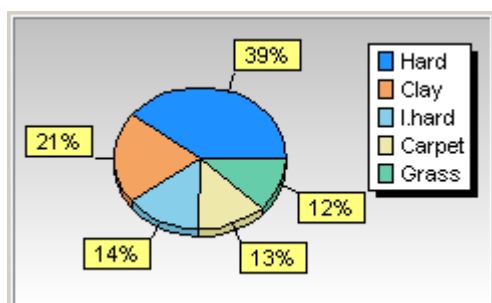


Obrázek 32 – Výsledek regresní analýzy pro závislost povrchu hrací plochy na výsledku. Provedeno shlukování a před ním jednoduchá kategorizace.

Vertikální osa rozdíl bodů hráčů rozdělený do intervalů po 200 bodech. Omezujeme velké množství vstupních hodnot na několik skupin. Záporný rozdíl značí, že vítěz měl původně nižší bodové hodnocení.

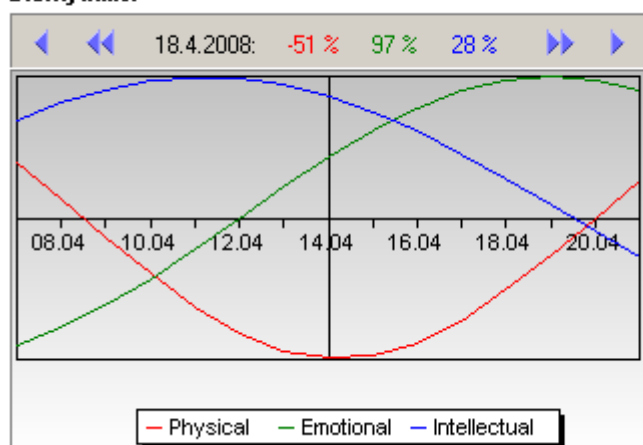
5. Budoucí rozšíření²⁶

Pro věrnější odhad pravděpodobností o výsledcích zápasů bych rád v budoucnu implementoval následující rozšíření, které budou s příslušnou vahou zahrnuty do sofistikovanějšího prediktivního algoritmu anebo uživatelům pomůžou se lépe zorientovat přehledným uspořádáním důležitých údajů a statistik.



Obrázek 33 ukazuje specializaci hráče pro různé povrchy hrací plochy.

Biorhythms:



Sign of zodiac:



Obrázek 34 biorytmy a horoskop©

ATP Profiles | WTA Profiles | Men's Pro Gear | Women's Pro Gear

Justine Henin		Ana Ivanovic	
	Racquet Wilson K Factor KTour Racquets		Racquet Yonex RQ IS 1 Tour Racquets
	Shoe Adidas Barricade V Women's Shoes White/Silver		Shoe Adidas Barricade IV Women's Shoes Whit/Silver
	Apparel Adidas		Apparel Adidas
Click any item to Order Now		Click any item to Order Now	

Obrázek 35 profily hráčů, jejich kariéra a statistiky, rozšíření i na čtyřhry a tenis žen

²⁶ Viz [29], [35], [36], [38]

Vyberte 2 hráče pro vzájemnou statistiku
 Berlocq, Carlos Proti Armando, Hugo

Tenis : vzájemná utkání Tento turnaj

Berlocq, Carlos **Armando, Hugo**

1 : 1

97 Startovní listina 205
 445 Body 192
 Argentina Země USA

Srovnání

Datum	Turnaj	Kolo	Kontinent	Povrch	Vítěz/ka	Výsledek	Výsledek
26/11-06	Naples, FL, USA	Semifinále	Amerika	Zelená antuka	Berlocq, Carlos	2:0	64 76
12/11-04	Copa Petrobras, Buenos Aires	Čtvrtfinále	Amerika	Antuka	Armando, Hugo	0:2	36 46

Poslední 5 Turnaje Berlocq Carlos

Datum	Kolo	Zápas	Výsledek	Sety
Miami USA, Amerika				
26/03-08	1.kolo	Berlocq, Carlos vs. Querrey Sam	0:2	67 36
Sunrise, Florida USA, Amerika				
20/03-08	1/8 finále	Berlocq, Carlos vs. Starace, Potito	0:2	46 26
19/03-08	3.kolo	Berlocq Carlos vs. Dlouhy, Lukas	2:1	46 76 63
PacificLifeOpen, Indian Wells, USA, Amerika				
15/03-08	1.kolo	Berlocq, Carlos vs. Simon, Gilles	0:2	26 36
TennisChannelOpen, Las Vegas, USA, Amerika				
04/03-08	3.kolo	Berlocq, Carlos vs. Starace, Potito	0:2	67 06
Abierto Mexicano Telcel Acapulco, Mexiko, Amerika				
26/02-08	3.kolo	Berlocq, Carlos vs. Canas, Guillermo	1:2	46 76 46

Poslední 5 Turnaje Armando, Hugo

Datum	Kolo	Zápas	Výsledek	Sety
Leon, Mexiko, Amerika				
27/03-08	1/8 finále	Armando, Hugo vs. Soeda, GO	1:2	46 63 57
25/03-08	3.kolo	Armando, Hugo vs. Juska, Andis	2:0	64 63
San Luis Potosi, Mexiko, Amerika				
19/03-08	1/8 finále	Armando, Hugo vs. Gonzalez, Santiago	0:2	67 36
18/03-08	3.kolo	Armando, Hugo vs. Mello, Ricardo	2:1	36 76 62
Salinas, Ekvádor, Amerika				
13/03-08	Čtvrtfinále	Armando, Hugo vs. Miranda, Ivan	1:2	46 61 06
12/03-08	1/8 finále	Armando, Hugo vs. Alves, Thiago	2:0	61 62
11/03-08	3.kolo	Armando, Hugo vs. Mayer, Leonardo	Vyhrál	76 20 (L.Mayer retired)
AustralianOpen Kvalifikace Melbourne, Austrálie, Austrálie				
09/01-08	1.kolo	Armando, Hugo vs. Amritraj Prakash	0:2	46 57
Puebla, Mexiko, Amerika				
20/11-07	3.kolo	Armando, Hugo vs. Gonzalez, Santiago	1:2	36 64 57

Obrázek 36 vzájemné statistiky, srovnání, poslední turnaje, alternativní zápasy

Speciálně profilované prostředí pro odhad vhodných sázek pro jednotlivé strategie, viz příslušná [kapitola](#) o strategiích jako valuebetting, asijské handicapy, sázení na kurzy 2-3 aj. Nejzajímavější se jeví systém 1-6 nula nebo vyhledávání sure bet na trhu sázkových kanceláří či využití bonusů a jejich výběr pomocí burzy sázek²⁷.

²⁷ Viz [41]

6 Závěr

Některé části diplomové práce jsem si předzpracoval v předmětech Získávání znalostí z databází a Soft computing. V rámci prvního předmětu jsem provedl hlavně analýzu dat. V rámci druhého předmětu návrh a implementaci GUI a některých algoritmů.

Projekt je možné ovládat pomocí GUI formulářů a dialogových oken propojených hlavní nabídkou. Na základním formuláři naleznete tabulku, která obsahuje data načtená z databáze, a to informace o hráčích, zápasech, chystaných utkáních, žebříčku hráčů aj. V dalším okně je možné vybrat zápas, nebo zadat vlastní zápas i s kurzy a učinit odhad.

Aplikace je schopná z internetu stáhnout seznam zápasů, které se teprve budou hrát a v okně odhadů predikovat vlastní kurzy na zápasy. Mělo by se jednat o tzv. správné kurzy s návratností 100%. Ty se pak dále dají využít při handicappingu, nebo jiných strategiích, jako je například valuebetting. Nejtěžší, a zatím ne příliš přesně pracující částí, je část kvalifikovaného odhadu. Na základě výsledků z předchozích let, na které se trénuje neuronová síť, se snažím využít natrénované neuronové sítě k predikci kurzů pro zápasy, které se budou teprve konat. Ale ze zkušeností, co jsem při studiu dalšího software, principů odhadů a handicappingu získal, je tento přístup jen část celkové mozaiky. Nejprestižnější sázkové kanceláře zaměstnávají týmy odborníků, využívají sofistikované komerční softwary s mnohem přesnějšími daty, statistikami, odhady a uvažují i aspekty jako je kausalita, motivace, kapacita a potenciál hráče, aktuální forma, linie prohraných x vyhraných zápasů a chování sázkařů. Hlavně chování sázkařů ovlivňuje dynamiku nabízených kurzů. Sázkové kanceláře experimentují s novými typy sázek, jako je například life betting a nebo velmi dlouhodobými sázkami, které je velmi těžké odhadnout. Jelikož známé teorie o předvídání budoucích jevů jsou převážně založeny na tom, že musíme co nejpřesněji rozumět minulosti a současnosti, abychom mohli odhadnout co se stane. Jestliže je ale událost ve vzdálenější budoucnosti²⁸, musíme další vývoj simulovat, ale to je tak náročné, viz butterfly effect, že to není ani technicky možné.

V dalších verzích aplikace by bylo zajímavé aplikovat sofistikovanější algoritmus pro analýzu kurzů, kde neuronové sítě budou jen jednou z částí. Bude zahrnovat vzájemné statistiky - porovnání z n posledních zápasů, specializace hráče, vliv povrchu, prestižnosti turnaje, získané tituly, biorytmy hráčů aj. Významnost těchto vlivů bude ve výsledku započítávána-oceněna různou vahou, bude tedy vhodné každý z nich parametrizovat. O správné nastavení vah parametrů se bude starat genetický algoritmus. Neustálé učení aplikace může být založeno na principu reinforcement learning.

²⁸ Viz [42]

Literatura

- [1] Půlpán Jaroslav, Dolování dat aneb Hledání skrytých souvislostí, <http://www.systemonline.cz/clanky/dolovani-dat-aneb-hledani-skrytych-souvislosti.htm> (duben '08), Zpravodajský portál časopisu IT Systems, PŘÍLOHA 6/2001
- [2] Řezanková Hana, Postupy používané při analýze dat, <http://badame.vse.cz/clanky/analyza-dat.php>, (duben '08), Projekt BADAME rok 2001
- [3] Ing. Molnár Karol, Úvod do problematiky umělých neuronových sítí, <http://www.elektrorevue.cz/clanky/00013/index.html>, (duben'08), ÚTKO FEI VUT
- [4] Marcel Jiřina, skriptum Neuronové sítě, CSLAB, FELK-ČVUT, 1995
- [5] Wikipedie, Otevřená encyklopedie, Matematická statistika, http://cs.wikipedia.org/wiki/Matematická_statistika, (duben'08),
- [6] Wikipedie, Otevřená encyklopedie Sázení po internetu, http://cs.wikipedia.org/wiki/Sázení_po_internetu, (duben'08)
- [7] Teorie sázení <http://www.sazeni.ws/teoriesazeni.html>, (duben '08)
- [8] Rejstřík sázkaře, <http://www.kurzovesazeni.com/rejstrik-sazkare.p8.html>, (duben '08)
- [9] Kurzové sázení <http://www.kurzovesazeni.com/>, (duben '08)
- [10] Jak tipovat a čeho se vyvarovat při sázení po internetu, <http://www.kurzovesazeni.com/jak-tipovat-uspesne.p6.html> (duben '08)
- [11] Sázení po internetu - krok za krokem: <http://www.kurzovesazeni.com/jak-zacit-krok-za-krokem.p3.html> (duben '08)
- [12] Systémy - strategie sázení, <http://www.kurzovesazeni.com/systemy-strategie-sazeni.p55.html>, (duben '08)
- [13] Recency Effect (Vliv nedávné minulosti), <http://www.kurzovesazeni.com/recency-effect-vliv-nedavne-minulosti.a38.html> (duben '08)
- [14] Rivalries - zápasy mezi tradičními rivaly, <http://www.kurzovesazeni.com/rivalries-zapasy-mez-tradicnimi-rivaly.a35.html> (duben '08)
- [15] Potenciál a kapacita v kurzovém sázení, <http://www.kurzovesazeni.com/potencial-a-kapacita-v-kurzovem-sazeni.a34.html> (duben '08)
- [16] Handicapping - analyzovat hru/zápas, <http://www.kurzovesazeni.com/handicapping-analyzovat-hruzapas.a33.html> (duben '08)
- [17] Sure bet (jistá výhra) – systém, <http://www.kurzovesazeni.com/sure-bet-jista-vyhra-system.a18.html> (duben '08)
- [18] Asijský handicap 0 : 1, <http://www.kurzovesazeni.com/asijsky-handicap-0-1.a25.html> (duben '08)

- [19] Asijský handicap 0 : $\frac{3}{4}$, <http://www.kurzovesazeni.com/asijsky-handicap-0-34.a24.html> (duben '08)
- [20] Asijský handicap 0 : $\frac{1}{2}$, <http://www.kurzovesazeni.com/asijsky-handicap-0-12.a23.html> (duben '08)
- [21] Asijský handicap 0 : $\frac{1}{4}$, <http://www.kurzovesazeni.com/asijsky-handicap-0-14.a22.html> (duben '08)
- [22] Asijský handicap 0 : 0, <http://www.kurzovesazeni.com/asijsky-handicap-0-0.a21.html>, (duben '08)
- [23] Systém 1 - 6 nula, <http://www.kurzovesazeni.com/system-1-6-nula.a20.html>, (duben '08)
- [24] Tip Info, Teorie sázení, <http://www.tipinfo.cz/?teorie-sazeni/>, (duben '08)
- [25] Value betting (sázka proti kurzu), <http://www.kurzovesazeni.com/value-betting-sazka-proti-kurzu.a19.html> (duben '08)
- [26] Sázení po internetu zdarma, <http://www.kurzovesazeni.com/sazeni-po-internetu-zdarma.p9.html> (duben '08)
- [27] Základy sázení <http://www.sazeni.ws/zakladysazeni.html>, (duben '08)
- [28] Sázení po internetu <http://www.sazeni-po-internetu.cz/jak-sazet-a-vyhraovat.aspx>, (duben '08)
- [29] Software OnCourt, <http://www.oncourt.info/>, (duben '08)
- [30] Data Files: All Competitions, Notes, ATP Men's Tour, Sony Ericsson WTA Women's Tour, <http://www.tennis-data.co.uk/alldata.php> (duben '08)
- [31] Bernacki Mariusz, Włodarczyk Przemysław, Principles of training multi-layer neural network using backpropagation algorithm http://galaxy.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html (duben '08), Electronics V, rok 2005
- [32] R. Rojas, Neural Networks - The Algorithm <http://page.mi.fu-berlin.de/rojas/neural/chapter/K7.pdf>, (duben '08), Springer-Verlag, Berlin, 1996
- [33] Fröhlich Jochen, Neural network with Java, <http://www.nnwj.de/backpropagation.html>, (duben '08), [Diploma thesis], Fachhochschule Regensburg, Department of Computer Science
- [34] Shereif S. El-Sheikh, Understanding and Implementing Neural Networks - The classic XOR problem, <http://gseacademic.harvard.edu/~elsheish/propagat.htm>, (duben '08)
- [35] Databáze a seznamy hráčů a turnajů s výsledky <http://www.stevegtennis.com/>, (duben '08)
- [36] Rankings, ATP, ATP Race Leaderboard, WTA, U.S. Open Series, No. 1s Guide, <http://www.sportsline.com/tennis/rankings>, (duben '08)
- [37] Šálek V., Použití neuronových sítí v loterii <http://www.cmail.cz/vsalek/neulot.htm>, (duben '08)

- [38] From Wikipedia, the free encyclopedia, Tennis <http://en.wikipedia.org/wiki/Tennis>, (duben '08)
- [39] Bc. Bělohávek Jiří, Využití neuronových sítí pro predikci sportovních výsledků, [Projekt do předmětu SFC], FIT VUTBR, 2007
- [40] Bc. Bělohávek Jiří, Bc. Šmerda Vojtěch, Dolování ve výsledcích tenisových utkání, [Projekt do předmětu ZZN], FIT VUTBR, 2007
- [41] Marek, Sázení na sport, <http://skalpuji.ic.cz/> (duben '08)
- [42] Eden Dan for Viewzone, Explore Your Future Consciousness <http://www.viewzone.com/future/index.html> (duben '08)

Seznam příloh

Příloha 1. Některé zajímavé zdrojové texty programu (proces learn, run v odděleném vlákně).

Příloha 2. CD s dokumentací, skripty, programem pro prostředí NetBeans IDE 6, zdroji, aj.

Příloha 1, zdrojové texty

```
168 void learn() {
169     if (error > minimumError && (super.learningCycle < super.maxLearningCycles
170         || super.maxLearningCycles == -1)) {
171         super.learningCycle++;
172         for (int i = 0; i <= lastPattern; i++) {
173             neuronLayerArray[0].setInput(inputPatternArray[i]);
174             for (int j = 1; j <= lastLayer; j++) {
175                 neuronLayerArray[j].computeInput(neuronLayerArray[j - 1],
176                     weightMatrixArray[j - 1]);
177                 neuronLayerArray[j].computeOutput();
178             }
179             neuronLayerArray[lastLayer].computeLayerError(targetPatternArray[i]);
180             layerOutputError[i] = neuronLayerArray[lastLayer].getLayerError();
181             for (int k = lastMatrix; k >= 0; k--) {
182                 weightMatrixArray[k].changeWeights(neuronLayerArray[k].getOutput(),
183                     neuronLayerArray[k + 1].getLayerError(), super.learningRate);
184                 if (k > 0) {
185                     neuronLayerArray[k].computeLayerError(neuronLayerArray[k + 1],
186                         weightMatrixArray[k]);
187                 }
188             }
189         }
190         double d = 0.0D;
191         for (int l = 0; l < layerOutputError.length; l++) {
192             for (int il = 0; il < layerOutputError[0].length; il++) {
193                 d += square(layerOutputError[l][il]);
194             }
195         }
196         error = Math.abs(d * 0.5D);
197         return;
198     } else {
199         super.stopLearning = true;
200         return;
201     }
202 }
```

Obrázek 37 BackpropagationNet.java, metoda learn, která pomocí backpropagation učí neuronovou síť

```

142  @Override
143  public void run() {
144      while (true) {
145          gui.EnableMenuTrain(false, "Train network: in progres..."); // train network
146          String[] args = {"100"};
147
148          System.out.println("\n\nNet structure:");
149          for (int i = 0; i < bpn.getNumberOfLayers(); i++) {
150              System.out.println("layer " + i + ": " + bpn.getNumberOfNeurons(i) + " neurons");
151          }
152          System.out.println("weights: " + bpn.getNumberOfWeights() + "\n\n");
153
154          // some optional method calls
155          bpn.setLearningRate(0.25);
156          bpn.setMinimumError(0.000005);
157          bpn.setMaxLearningCycles(100); // -1 = no maximum cycle
158          bpn.setDisplayStep(Integer.parseInt(args[0])); // get the argument
159
160          System.out.print("\n\nrallright, let's learn...\n\n");
161          bpn.resetTime();
162          while (!bpn.finishedLearning()) { // while the net learns
163              bpn.learn(); // perform one learning step
164              if (bpn.displayNow()) {
165                  output();
166              } // display the current results
167          }
168          output();
169          System.out.println("\n\nrFINISHED.");
170          gui.EnableMenuTrain(true, "Train network: finished");
171          suspend();
172          bpn.stopLearning = false;
173          bpn.learningCycle = 0;
174      }
175  }

```

Obrázek 38 BPN.java, smyčka běžící v odděleném vlákně, trénuje neuronovou síť