# Návrh riešenia

V našej práci sme pracovali s dátami zo symetrickej hry Poker - *No Limit Texas Holdem*. Dáta boli získané od reálneho hráča z herne Pokerstars. V analýze sme opísali a identifikovali základné štatistické a taktické faktory hry, ako aj prístup, pomocou ktorého sme predikovali nami stanovené výsledky priebehu hry, ktorých určenie sme si zvolili ako cieľ našej práce. Konečný výber použitých faktorov spolu s odôvodnením výberu je poskytnutý v častiach práce *Implementácia predspracovania dát pre predikciu víťaza hry* a *Implementácia predspracovania dát pre určenie pravdepodobností kartových kombináci*. V tejto časti sme poskytli aj náhľad na získavanie informácií a určovanie nami stanoveného výsledku hry a jeho využiteľnosť.

V našej práci sme sa rozhodli zamerať sa na dva rôzne typy predikcie. Vstupné informácie pre predikciu oboch výsledkov majú rovnaký charakter. Prvý experiment bol zameraný na určovanie víťaza jednotlivých hier. Túto časť sme si určili za čiastočný cieľ, ktorý nám poskytol náhľad na overenie využitia vstupných veličín využívaných pre predikciu. Hlavným cieľom našej práce bolo určenie pravdepodobností typov kartových kombinácií pre protihráča pre konkrétnu hru. Ďalšia časť tejto kapitoly je preto zameraná na návrh riešenia určený pre dosiahnutie uvedeného hlavného cieľa.

Našu prácu sme rozdelili do dvoch častí. V prvej časti sme sa zamerali na predspracovanie dát, z ktorých sme dolovali štatistické informácie o hráčoch potrebné na zaradenie hráčov do zhlukov. Na základe týchto údajov sme vytvárali zhluky hráčov s podobným správaním. V druhej časti sme pracovali s rovnakými dátami a však ich spracovanie sme smerovali na popis priebehu hry. Pomocou rozdelenia priebehu jednotlivých hier na akcie vykonávané hráčmi sme získali dostatočnú množinu dát na predikovanie typu kartovej kombinácie hráča. Pre túto časť našej práce sme sa po dôkladnej analýze rozhodli využiť rekurentnú neurónovú sieť s dlhou krátkodobou pamäťou. Na základe vstupov, ktoré reprezentujú priebeh jednotlivých hier prostredníctvom vektorov akcií sme získali pomocou špecifických modelov pre jednotlivých hráčov, resp. skupiny hráčov výstupy, ktoré reprezentovali pravdepodobnosti typov kartovej kombinácie hráča.

## Nespracované dáta

Priebeh jednotlivých hier bol zapísaný prostredníctvom detailných údajov o hre, ktoré vznikli použitím nástroja Holdem Manager 2 a bolo možné ich dolovať priamo z databázy, respektíve z vytvorených textových súborov. Údaje majú podobu voľne štruktúrovaného textu.

V našej práci sme sa zaoberali analýzou dostupných dát získaných od profesionálneho hráča online pokru z herne Pokerstars. Dáta, ktorými sme disponovali obsahovali priebeh 273 tisíc hier a spolu sa v nich vyskytovalo viac ako 12 tisíc hráčov. Dáta popisovali priebeh pokrovej hry *No Limit Texas Holdem[[1]](#footnote-1)*. Priebeh hry je popísaný pomocou inicializačných dát (id hry, dátum a čas hry, pozícia hráčov vzhľadom na pozíciu button, vkladanie slepých stávok – small blind a big blind) a podrobný popis jednotlivých častí hry, skladajúcich sa z akcií, stavov v ktorých sa priebeh hry nachádza (inicializácia, preflop, flop, turn, river). V každom stave boli chronologicky zaznamenané akcie všetkých hráčov, ktorí sa do daného stavu dostali.

Príklad jednej pokrovej ruky (hry):

PokerStars Hand #146448977997: Hold'em No Limit ($0.50/$1.00 USD) - 2016/01/04 15:40:08 CET [2016/01/04 9:40:08 ET]

Table 'Pretoria' 9-max Seat #9 is the button

Seat 2: ilaviiitech ($100 in chips)

Seat 3: matiko09 ($101.50 in chips)

Seat 4: ardo888 ($101 in chips)

Seat 6: Dachish ($111.50 in chips)

Seat 7: Momess ($128.09 in chips)

Seat 9: finito.com ($52.67 in chips)

ilaviiitech: posts small blind $0.50

matiko09: posts big blind $1

\*\*\* HOLE CARDS \*\*\*

Dealt to matiko09 [4h 8c]

ardo888: folds

Dachish: folds

Momess: folds

finito.com: raises $1 to $2

ilaviiitech: calls $1.50

matiko09: folds

\*\*\* FLOP \*\*\* [Ks 4s Kh]

ilaviiitech: checks

finito.com: checks

\*\*\* TURN \*\*\* [Ks 4s Kh] [5c]

ilaviiitech: bets $2.86

finito.com: calls $2.86

\*\*\* RIVER \*\*\* [Ks 4s Kh 5c] [Kc]

ilaviiitech: bets $6.14

finito.com: raises $17.14 to $23.28

ilaviiitech: raises $71.86 to $95.14 and is all-in

finito.com: calls $24.53 and is all-in

Uncalled bet ($47.33) returned to ilaviiitech

\*\*\* SHOW DOWN \*\*\*

ilaviiitech: shows [Tc Td] (a full house, Kings full of Tens)

finito.com: shows [Jd Kd] (four of a kind, Kings)

finito.com collected $103.54 from pot

\*\*\* SUMMARY \*\*\*

Total pot $106.34 | Rake $2.80

Board [Ks 4s Kh 5c Kc]

Seat 2: ilaviiitech (small blind) showed [Tc Td] and lost with a full house, Kings full of Tens

Seat 3: matiko09 (big blind) folded before Flop

Seat 4: ardo888 folded before Flop (didn't bet)

Seat 6: Dachish folded before Flop (didn't bet)

Seat 7: Momess folded before Flop (didn't bet)

Seat 9: finito.com (button) showed [Jd Kd] and won ($103.54) with four of a kind, Kings

## Návrh štruktúry spracovania informácií z jednotlivých hier

Pre našu prácu bolo dôležité dôsledné navrhnutie štruktúry spracovania hier. Bolo potrebné správne spracovanie a uchovanie veľkého množstva detailov, ktoré boli potrebné pre získanie dostatočnej množiny informácii pre tvorbu zhlukov hráčov a pre predikciu typu kartovej kombinácie hráča. Z analýzy štruktúry hry, taktických a technických aspektov hry sme získali dostatočné množstvo poznatkov, ktoré tvorili dostačujúci základ pre vytvorenie štruktúry spracovania. V tejto kapitole sú popísané základné entity, ktoré sme sa rozhodli využívať pre štruktúru spracovania hry a väzby medzi entitami.

Pre každú hru sme uchovávali inicializačné údaje (hodnota malého blindu a veľkého blindu), ktoré nám slúžili na zjednotenie hier pri rozdielnych herných limitoch pre možnosť využitia všetkých hier. Pre každú hru sme uchovávali zoznam hráčov radených podľa pozícii na stole, zoznam akcií z ktorých sa jednotlivé hry skladali a prvé tri spoločné karty vyložené na stole.

Jednotlivé hry sme sa rozhodli rozdeliť na akcie. Každá akcia reprezentovala rozhodnutie hráča v priebehu hry. Pre ďalšie využitie akcie bolo potrebné získanie a uchovanie detailov z priebehu hry. Na opis akcie sme využívali a uchovávali informácie:

* Hráč – entita hráča, ktorý vykonáva akciu. Každý hráč si uchováva hodnoty svojho stacku pre všetky hry a odhalené kartové kombinácie pre všetky hry.
* Typ rozhodnutia – rozhodnutia hráčov sme sa rozhodli kategorizovať do piatich základných kategórii (staviť, zvýšiť, dorovnať, držať akciu, zložiť).
* Stav hry – hra sa počas jej priebehu dostáva do piatich stavov, ktoré sme sa rozhodli využiť pre kategorické rozdelenie stavu hry (inicializácia hry, preflop, flop, turn, river).
* Hodnota stávky – vyjadruje hodnotu stávky hráča, ktorý vykonáva akciu.
* Stack – vyjadruje hodnotu hráčových žetónov, ktoré môže v danej hre použiť.

## Reprezentovanie získaných hodnôt

Pre využitie získaných informácií bola dôležitá správna reprezentácia údajov. Na základe získaných poznatkov sme sa rozhodli veličiny normalizovať, resp. vykonať úpravy, ktoré nám zabezpečili jednoduché, nami požadované využitie veličín počas celého priebehu chodu programu bez potreby ďalšej transformácie.

### Normalizovaná hodnota stávky

Jednotlivé hry nie sú odohrané na jednej úrovni výšky inicializačných stávok hry. Pre získané hodnoty stávok sme využili normalizáciu v závislosti na hodnote veľkého blindu, inicializačnej stávky pre hru. Jednotlivé hodnoty sme prevádzali na počet veľkých blindov. Táto normalizácia je opísaná vzorcom:

### Reprezentácia karty

Na základe získaných poznatkov z analýzy uvedených v časti práce *Súvisiace práce z oblasti využitia neurónových sietí spojených s adaptovaním na symetrickú hru poker* sme sa rozhodli kartu, ktorá obsahuje údaje hodnota a znak reprezentovať vektorom dĺžky 17. Vektor bol zložený z dvoch častí. Prvá časť, prvých 13 miest vektora bolo vyhradených pre reprezentáciu hodnoty karty. Hodnoty sú zoradené v postupnosti A,2,3,4,5,6,7,8,9,T,J,Q,K. Druhá časť, miesta vektoru 14 – 17 reprezentovali znak karty. Znaky boli zoradené v postupnosti s,h,d,c. Pre reprezentáciu jednej karty mohli hodnoty vektora nadobúdať hodnoty 0, resp. 0,3, ak bola nájdená daná hodnota alebo znak. V prípade reprezentácie kartovej kombinácie alebo vyložených spoločných kariet sme vytvárali súčet vektorov.

### Kategorizácia typov kartových kombinácii

Každý hráč hrá počas jednej hry s jednou kartovou kombináciou zloženou z dvoch vlastných súkromných kariet. Na základe informácií o veľkom počte možných kartových kombinácii, uvedených v časti práce *Modelovanie protihráča* sme sa pre nedostatok opakujúcich sa identických situácii v histórii hier jedného hráča rozhodli pre reprezentáciu hráčových kariet využiť základnú kategorizáciu typov kartových kombinácii v súčinnosti s vyloženými verejnými kartami. Pre rozhodovanie hráča v stave hry flop nie je podstatná presná kartová kombinácia protihráča, postačujúcim údajom je aj kategória do ktorej sú protihráčove karty v súčinnosti s vyloženými verejnými kartami zaradené. Základné informácie pre zaraďovanie kartových kombinácii do jednotlivých kategórii sme čerpali z poznatkov uvedených v časti práce *Základná kategorizácia rúk*. Kategórie s pravidlami zaraďovania:

* Hotová ruka – hráčove karty v súčinnosti s verejnými kartami tvoria minimálne jeden pár alebo silnejšiu kombináciu. Z tejto kategórie sme sa rozhodli vyňať párové kombinácie hráča tvorené výlučne zo súkromných kariet hráča, ak je párová kombinácia nižšej hodnoty ako najnižšia vyložená spoločná karta. V prípade vyňatia kartovej kombinácie radíme túto kombináciu do kategórie „prázdna ruka“.
* Nábeh na postupku - hráčove karty v súčinnosti s verejnými kartami obsahujú presne štyri hodnoty z po sebe nasledujúcich piatich hodnôt zo základnej postupnosti kariet.
* Nábeh na farbu - hráčove karty v súčinnosti s verejnými kartami obsahujú presne štyri karty s rovnakým znakom.
* Nábeh na postupku a farbu – hráčove karty v súčinnosti s verejnými kartami spĺňajú podmienky pre zaradenie do kategórii „Nábeh na farbu“ a „Nábeh na postupku“.
* Hotová ruka s nábehom - hráčove karty v súčinnosti s verejnými kartami spĺňajú podmienky pre zaradenie do kategórii „Hotová ruka“ a „Nábeh na postupku“ alebo „Hotová ruka“ a „Nábeh na farbu“.
* Prázdna ruka - hráčove karty v súčinnosti s verejnými kartami nespadajú do žiadnej z vyššie uvedených kategórii.

## Rozdelenie hráčov

Pre správne pristupovanie k vyhodnocovaniu rozhodnutí hráčov je dôležité zohľadňovať rôzne štýly hry hráčov. Na základe informácii uvedených v časti práce *Základné rozdelenie hráčov* a *Modelovanie protihráča* je možné zaradenie jednotlivých hráčov do skupín na základe štatistických ukazovateľov získaných z odohratých hier. V našej práci sme rozdelili hráčov na profesionálnych, amatérskych a hráčov nezaradených, s ktorými sme ďalej nepracovali.

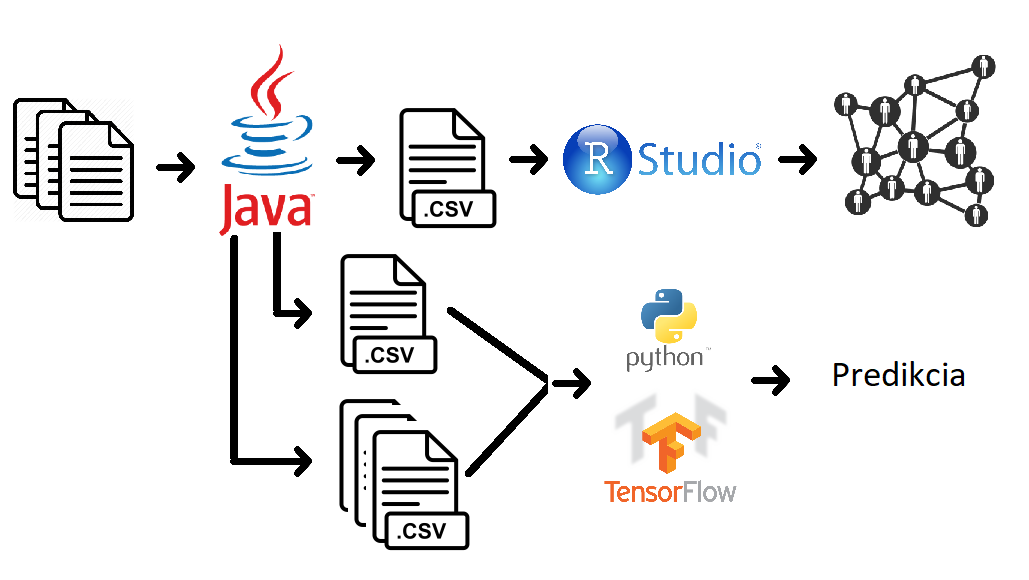
Nezaradení hráči boli tí, ktorých sme pre nedostatočné množstvo dát nedokázali zaradiť medzi profesionálov, ale ich hra im prináša zisk. Vychádzali sme z poznatku, že väčšina amatérskych hráčov svojou hrou nedosahuje zisk. Do tejto skupiny sme zaradili tiež hráčov, z ktorých hry máme takú malú vzorku, že by v našich dátach spôsobovali zbytočný, nechcený šum.

Pre profesionálov sme mali dostatočnú históriu odohratých rúk, a preto sme mohli zvoliť individuálny prístup pri tvorbe modelu pre predikciu typov kartových kombinácií. Takýto model je špecifický, a mal by sa dokázať vysporiadať so špecifikami jednotlivého hráča, ktoré by pri začlenení medzi ostatných hráčov zanikli, čím by sme odbúrali možnosť správnej kategorizácie kartovej kombinácie na základe priebehu hry.

Amatérski hráči nemali dostatočnú históriu pre individuálny prístup, a preto sme potrebovali nájsť skupiny hráčov s podobným správaním v priebehu hry. Rozhodli sme sa nevytvárať naše skupiny klasickým prístupom. Nezaraďovali sme hráčov na základe štatistických ukazovateľov získaných z priebehu hry do vopred vytvorených skupín spomenutých v častiach práce *Základné rozdelenie hráčov* a *Modelovanie* . Takéto striktné zaraďovanie môže zapríčiniť, že hráči v jednej skupine budú mať výrazne rozdielne správanie. Pre našu prácu sme zvolili prístup vytvárania zhlukov hráčov, kedy sme pre zaradenie poskytli štatistické faktory hráčov, avšak pre samotné zaraďovanie hráčov do zhlukov sme využili známy algoritmus pre hierarchické zhlukovanie, ktorému upravíme parametre podľa našich potrieb.

## Koncept riešenia

Na Obrázok 14 môžeme vidieť koncept riešenia. Spracovanie a príprava dát pre vytváranie zhlukov a pre predikciu bola napísaná v programátorskom jazyku Java. Prostredníctvom .csv súboru sme poskytli všetky potrebné informácie pre vytvorenie zhlukov hráčov v nástroji Rstudio. Pre predikciu pravdepodobností typov kartových kombinácií sme vytvorili prostredníctvom jazyka Java osobitne pre každého hráča, resp. skupinu hráčov jeden vstupný .csv súbor pre každú hru, reprezentujúci priebeh hry prostredníctvom informácií o jednotlivých akciách hry, v ktorej boli hráčove karty odhalené. Pre výstup sme vytvorili jeden .csv súbor pokrývajúci všetky hry. Súbor obsahoval typ kartovej kombinácie hráča pre každú hru. Súbory určené pre predikciu sme spracovali v jazyku Python. Pre vytvorenie modelu a predikcie sme využili TensorFlow a ďalšie pomocné knižnice.



Obrázok 14: Koncept riešenia opisujúci tok dát. Tvorba zhlukov – dáta popisujúce priebeh hry -> predspracovanie dát, tvorba štatistík -> dáta so štatistikami pre hráčov -> tvorba zhlukov -> zhluky. Predikcia kartovej kombinácie – dáta popisujúce priebeh hry -> predspracovanie dát - vytvorenie súboru s typom kartovej kombinácie, ktorú drží hráč na ruke v danej hre, vytvorenie súborov s opisom jednotlivých hier prostredníctvom akcii -> spracovanie dát + predikčný model -> výsledky trénovania, validácie a testovania.

# Spracovanie dát

## Technická dokumentácia

V tejto časti uvádzame technickú dokumentáciu pre vytvorený prototyp pre spracovanie dát s cieľom opisu jednotlivých hier prostredníctvom akcií. Prostredníctvom návrhu a implementácie tohto prototypu boli pokryté nasledujúce prípady použitia, ktoré sme si stanovili pri špecifikácii systému.

*UC03. Spracuj surové dáta pre opis hráčov pomocou štatistických faktorov* – prostredníctvom nami navrhnutej a implementovanej metódy *processSourceForInit*(), ktorá je súčasťou prototypu riešenia dokážeme spracovať surové dáta do štruktúry reprezentujúcej jednu hru. Zjednodušený model je načrtnutý v diagrame na Obrázok 24.

*UC04. Rozdeľ hráčov na profesionálnych a amatérskych*  – pomocou nami definovaných pravidiel sme rozdelili hráčov na profesionálnych, ktorý majú odohratých viac ako 15000 hier a amatérskych pre ktorých sme vytvorili pravidlá zaraďovania do tejto skupiny.

*UC05. Vytvor vstupné dáta pre zhlukovanie amatérskych hráčov so štatistickými údajmi o hráčoch*  – pomocou nami navrhnutých metód na výpočet jednotlivých štatistických faktorov sme určili hodnoty a vytvorili súbor so štatistickými faktormi pre amatérskych hráčov.

*UC06. Spracuj surové dáta pre opis akcií pre určovanie pravdepodobností kategórií kartových kombinácií* – prostredníctvom nami navrhnutej a implementovanej metódy *processSourceForPrediction*(), ktorá je súčasťou prototypu riešenia dokážeme spracovať surové dáta do štruktúry reprezentujúcej jednu hru. Zjednodušený model je načrtnutý v diagrame na Obrázok 24.

*UC07. Vytvor vstupné a výstupné dáta pre predikciu pravdepodobností kategórií kartových kombinácií* – prostredníctvom nami navrhnutej a implementovanej metódy spracovania, ktorá je súčasťou prototypu riešenia dokážeme spracovať načítané dáta a vytvoriť vhodné súbory pre použitie neurónovou sieťou.

## Implementácia prototypu

Prototyp Holdem Data Creator je vytvorený v programátorskom jazyku Java v prostredí Eclipse. Prototyp spracúva surové dáta za účelom vytvorenia vstupných súborov pre Players Clusterer a pre Holdem CCPredictor. Základná štruktúra spracovania dát je opísaná diagramom na Obrázok 24.

Program generuje priečinky, do ktorých ukladá súbory s dátami:

* Priečinok generuje iba ak nie je vytvorený:

Data/PokerPredictor/

Data/PokerPredictor/HandToPredict/ - predpripravený na uloženie ruky na predikciu

Data/PokerPredictor/ModelSaver/ - predpripravený na ukladanie modelov

Data/PokerPredictor/ModelSaverEmpty/ - predpripravený na uloženie základného modelu

Data/PokerPredictor/ModelsPlots/ - predpripravený na ukladanie grafov presnosti modelov vytváraných pri trénovaní, validácii a testovaní modelov

* Priečinok generuje vždy:

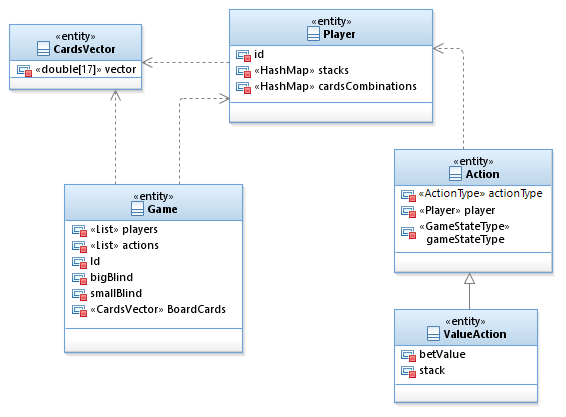
Data/PokerPredictor/Clustering/ - vstupné dáta pre tvorbu zhlukov

Data/PokerPredictor/ClusteringGroups/ - predpripravené pre uloženie zhlukov

Data/PokerPredictor/Prediction/ - dáta profesionálnych hráčov pre predikciu

Data/PokerPredictor/PredictionGroups/ - dáta zhlukov hráčov pre predikciu

Jednotlivé cesty sú uložené v konfiguračnom súbore config.java.



Obrázok 24: Zjednodušený model štruktúry spracovania dát pre opis hry.

## Inštalačná príručka

Spustenie prototypu je možné na počítači s operačným systémom Windows. Pre spustenie prototypu je potrebné vykonanie nasledujúcich úkonov:

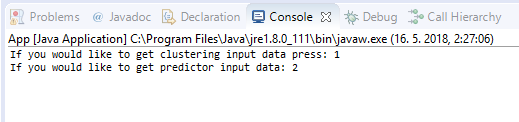
* Inštalácia Java (JDK) verzia 1.8.
* Inštalácia prostredia Eclipse.
* Umiestnenie dát s pokrovými hrami určenými na spracovanie do priečinku:

Data/PokerPredictor/HandHistory/

* Formát súborov s dátami – štandardizovaný formát spracovania Holdem Managerom 2.
* Import projektu Holdem Data Creator, ktorý je súčasťou priloženého dátového disku do prostredia Eclipse. Po importe je projekt možné spustiť.

## Používateľská príručka

Po otvorení programu v prostredí Eclipse je možné program spustiť (trieda App.java). Po spustení programu sa v konzole objavia možnosti spustenia (viď Obrázok 25). Pre vytvorenie dát pre zhlukovanie stlačte klávesu „1“ a potvrďte možnosť klávesou „enter“. Pre vytvorenie dát pre predikciu kategórií kartových kombinácií stlačte klávesu „2“ a potvrďte možnosť klávesou „enter“. Pre druhú možnosť musia byť súbory so zhlukmi hráčov uložené v priečinku C:/PokerPredictor/ClusteringGroups. Po vygenerovaní súborov sa program automaticky ukončí.



Obrázok 25: Možnosti spustenia programu Holdem Data Creator.

1. http://www.pokerlistings.com/poker-rules-texas-holdem [↑](#footnote-ref-1)