

Charles University
Faculty of Social Sciences
Institute of Economic Studies



MASTER'S THESIS

**Non-Linear Classification as a Tool for
Predicting Tennis Matches**

Author: **Bc. Jakub Hostačný**

Supervisor: **RNDr. Matúš Baniar**

Academic Year: **2017/2018**

Abstrakt

V tejto diplomovej práci skúmame predikčnú presnosť a výkon pri stávkovaní u štyroch strojovo učiacich sa algoritmov - penalizovaná logistická regresia, náhodný les, posilnené stromy a neurónové siete. Pri práci využívame 40 310 ATP zápasov hraných počas obdobia 1/2001-10/2016. Čo sa týka predikčnej presnosti, naše modely prekonávajú najlepšie modely súčasnosti pre predikovanie negrandslamových (69%) ako aj modely pre predikovanie grandslamových zápasov (79%). Všetky špecifikácie modelov sú presnejšie ako predikovanie na základe rebríčkového postavenia hráčov, zatiaľ čo väčšina špecifikácií je presnejších ako predikovanie na základe vypísaných kurzov stávkových kancelárií. Čo sa týka návratnosti pri stávkovaní, vytvorili sme šesť profitabilných stratégií pre stávkovanie na favoritov pre negrandslamové zápasy (návranosť investície v rozmedzí 0.8-6.5%). Taktiež sme identifikovali desať profitabilných stratégií pre stávkovanie na favoritov pre grandslamové zápasy (návranosť investície v rozmedzí 0.7-9.3%). Naše modely prinášajú vyššiu návratnosť ako stávkovanie na rebríčkového či kurzového favorita. Neurónové siete a náhodné stromy prinášajú najvyšší celkový zisk, zatiaľ čo posilnené stromy vykazujú najvyššiu percentuálnu návratnosť. Výsledky ďalej ukazujú, že veľkosť stávky závislá na Kellyho kritériu je optimálnejšia ako konštantná stávka pre stávkovanie na favoritov.

Klasifikace JEL

C01, C38, C45, C51, C52, C53, C55

Klíčová slova

neurónové siete, logistická regresia, náhodný les, posilnené stromy, tenisové predpovede, stávkovanie na tenis, tenisové modelovanie

E-mail autora

jakubhstn@gmail.com

E-mail vedoucího práce

matus.baniar@gmail.com