

# NAPOVED MENJALNEGA TEČAJA AMERIŠKEGA DOLARJA NA PODLAGI MENJALNIH TEČAJEV TUJIH VALUT

*Valentin Koblar*  
Univerza v Novi Gorici  
Poslovno-tehniška fakulteta  
e-mail: [valentin.koblar@kolektor.si](mailto:valentin.koblar@kolektor.si)

## POVZETEK

V poročilu je predstavljen pristop, k modeliranju oziroma odkrivanju povezav med spremembo vrednosti tečaja ameriškega dolarja in ostalimi tujimi valutami. Uporabljeni sta dve različni metodi za odkrivanje povezav med tečaji valut in sicer linearna regresija in modelno drevo (model tree). Za obe metodi so predstavljeni vhodni podatki in rezultati metod. Za lažjo predstavbo so v poročilo dodani tudi grafikoni z rezultati.

## 1 UVOD

Naloga, ki sem si jo zastavil pri predmetu »*Odkrivanje zakonitosti v tabelarnih podatkih*« je, ali lahko na podlagi podatkov tujih valut napovemo tečaj, oziroma spremembe vrednosti tečaja ameriškega dolarja (v nadaljevanju USD). Torej odkriti je bilo potrebno ali lahko dejansko na podlagi ostalih tečajev, karkoli sklepamo o vrednosti tečaja USD. V ta namen je bilo seveda potrebno najprej pridobiti ustrezne podatke, jih spraviti v ustrezno obliko za obdelavo v WEKI in jih potem s tem programskim orodjem tudi obdelati. Rezultat naj bi bil model, ki nam omogoča napoved vrednosti tečaja USD ali pa vsaj pričakovano smer gibanja vrednosti tega tečaja.

## 2 PRIDOBIVANJE IN PRIPRAVA PODATKOV ZA OBDELAVO

Podatke za tečaje valut sem pridobil na spletni strani »Banke Slovenije«<sup>1</sup>. Na tej spletni strani lahko pogledamo tečaje v preteklosti za različne valute in jih tudi shranimo v obliki tabelarnih podatkov (txt datoteka).

Seveda tako pridobljeni podatki niso direktno uporabni za obdelavo v WEKI, ampak jih je potrebno najprej ustrezno pripraviti in obdelati. Sam sem kot za vhodne podatke izbral podatke vrednosti tečaja 15 različnih tujih valut za 5 preteklih let (od 1.1.2002 do 31.12.2006). Podatke sem nato obdelal v programu MS Excel.

Najprej sem te podatke moral spraviti v pravo obliko (podatki, ki jih dobimo na spletni strani Banke Slovenije so sortirani po datumu za vseh 15 tujih valut) in združiti skupaj vse podatke preteklih 5 let (podatke dobimo za

vsako leto posebej). Pri tem se je seveda potrebno zavedati, kakšen je naš cilj. Kot ciljno spremenljivko bomo vzeli spremembo tečaja USD in ne samo vrednost tečaja. Kot vrednost menjalnega tečaja valut sem vzel srednji menjalni tečaj. Vse vrednosti in spremembe tečaja so še v enotah SIT (nisem pretvarjal v evre).

Valute, katerih vrednost menjalnih tečajev predstavljajo vhodne podatke so:

AUD	– avstralski dolar
CAD	– kanadski dolar
CHF	– švicarski frank
CZK	– češka krona
DKK	– danska krona
EUR	– evro
GBP	– britanski funt
HRK	– hrvaška kuna
HUF	– madžarski forint
JPY	– japonski jen
NOK	– norveška krona
PLN	– poljski zlot
SEK	– švedska krona

Kot že rečeno zanimajo nas predvsem spremembe tečaja, zato sem v MS Excelu izračunal spremembe tečajev vseh valut na podlagi tedenskega zamika. Tako obdelane podatke, je bilo potrebno samo še shraniti v WEKI razumljivo obliko (tj. csv oblika datoteke – uporaba "," in ".").

## 3 OBDELAVA PODATKOV V WEKI

V primeru vrednosti sprememb tečajev valut, gre za zvezno ciljno spremenljivko in ne diskretno. Ker je ciljna spremenljivka definirana moramo uporabiti napovedno indukcijo. WEKA nam omogoča uporabo različnih metodologij za izdelavo napovednega algoritma, sam pa sem preizkusil dve, in sicer:

- LeastMedSq – linearna regresija
- M5P – model tree

V vseh primerih sem za validacijo izračunanega modela uporabil prečno preverjanje.

<sup>1</sup> <http://www.bsi.si/>

Kot že rečeno me je pri modeliranju zanimalo predvsem sprememba tečaja USD. Da bi lahko različne modele med seboj objektivno primerjal, sem pri vsakem modelu izračunal standardni odklon razlike dejanskih in izračunanih sprememb tečajev (pri izračunu sem zajel vse podatke o spremembi tečaja USD).

Enačbi za izračun standardnega odklona sta naslednji:

$$\Delta_{RAZ} = \Delta_{USD} - \Delta_{NAP} \quad [1]$$

$\Delta_{RAZ}$  - razlika med napovedjo in izračunano vrednostjo iz modela

$\Delta_{USD}$  - dejanska sprememba tečaja USD v enem tednu

$\Delta_{NAP}$  - izračunana sprememba tečaja USD v enem tednu

$$st\_dev = \sqrt{\frac{(\Delta_{RAZ} - \bar{\Delta}_{RAZ})^2}{(n-1)}} \quad [2]$$

$st\_dev$  - standardni odklon razlike med dejansko in izračunano vrednostjo tečaja USD

$\Delta_{RAZ}$  - razlika med napovedjo in izračunano vrednostjo iz modela

$\bar{\Delta}_{RAZ}$  - povprečna razlika med napovedjo in izračunano vrednostjo iz modela

$n$  - število vseh zajetih podatkov

Vsi grafikoni v tem poročilu so narisani na podlagi sprememb tečaja 50 zaporednih tednov (50 tednov zaradi preglednosti) in sicer prvih 50 zaporednih tednov v podatkih.

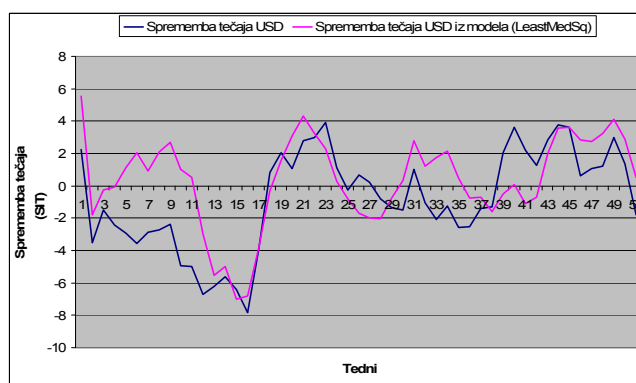
### 3.1 Linearna regresija in rezultat modela (algoritem LeastMedSq)

Torej ena od možnosti za določanje oziroma izračun modela za napoved vrednosti tečaja USD je linearna regresija. V teoriji sam model ni najboljši, saj v primeru velikih in nenadnih skokov model ne sledi posameznim vrednostim, pač pa le minimizira povprečje kvadratov odstopanj (ista linearna premica čez vse podatke). V našem primeru ko izberemo kot vhodni podatek vrednosti menjalnih tečajev vseh 14 valut nam WEKA poda naslednji model:

$$USD = \quad [3]$$

$$\begin{aligned} & -0,1216 * AUD + \\ & 0,7047 * CAD + \\ & -0,4289 * CHF + \\ & -6,2987 * CZK + \\ & -10,7899 * DKK + \\ & 0,1079 * EUR + \\ & 0,2449 * GBP + \\ & 1,46 * HRK + \\ & -7,5528 * HUF + \\ & 24,4295 * JPY + \\ & -0,0586 * NOK + \\ & 0,369 * PLN + \\ & -0,0997 * SEK + \\ & -3,1899 * SKK + \\ & 0,1766 \end{aligned}$$

Če v praksi preizkusimo ta model (LeastMedSq) nam poda naslednji rezultat (na grafikonu 1 je zaradi nazornosti prikazano samo 50 vrednosti sprememb tečaja USD in napoved tečaja iz izračunanega modela):



Grafikon 1: Sprememba tečaja USD (dejanski in napovedani)

Kot lahko vidimo iz zgornjega grafikona, model približno izračunava vrednost spremembe menjalnega tečaja USD, vendar se pojavi kar veliko odstopanje. Standardni odklon razlike dejanskih in izračunanih sprememb tečajev znaša 2,2230. Vsekakor je zanimivo spoznanje, da dejansko neka povezava med menjalnimi tečaji obstaja, težko pa jo je čisto natančno določiti (vsaj s tako enostavnimi algoritmi kot je linearna regresija). Iz tega razloga sem poskušal izgraditi model še z algoritmom »modelnega drevesa« (model tree).

### 3.2 Modelno drevo in rezultat modela (algoritem M5P)

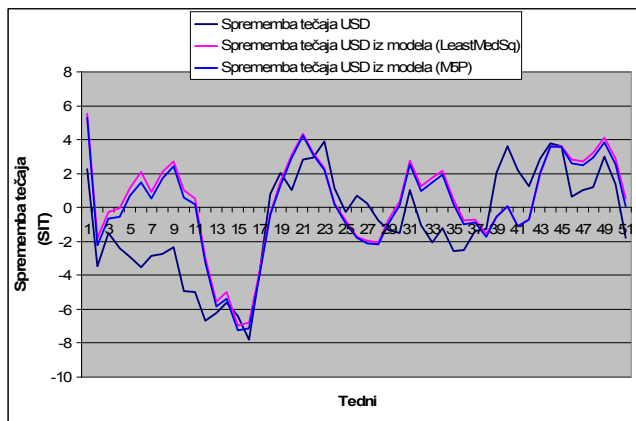
V teoriji naj bi bila boljša metoda kot linearna regresija »modelno drevo«. Razlika med tema dvema metodama je v tem, da ima »modelno drevo« v vsakem »listu« (vsaki skupini podatkov) linearno regresijo – torej linearna regresija se prilagaja posamezni skupini podatkov in ni skupna za vse podatke. Če v WEKI nastavimo parameter *unpruned* pri M5P algoritmu *false* dobimo naslednji rezultat:

$$USD = \quad [4]$$

$$\begin{aligned} & -0,1305 * AUD \\ & + 0,7223 * CAD \\ & - 0,3494 * CHF \\ & - 5,7664 * CZK \\ & - 6,9742 * DKK \\ & + 0,2662 * GBP \\ & + 1,393 * HRK \\ & - 9,3184 * HUF \\ & + 22,5691 * JPY \\ & + 0,3487 * PLN \\ & - 2,7299 * SKK \\ & + 0,1675 \end{aligned}$$

Kot vidimo se rezultat razlikuje od prejšnjega (linearna regresija), kar je posledic drugačnega algoritma. Ta rezultat v bistvu predstavlja samo en list modelnega drevesa, za vse primere iz podatkov (primer sem naredil zaradi lažje primerjave zgrajenih modelov). Praktični izračuna vrednosti

sprememb menjalnega tečaja nam kaže grafikon 2 (zaradi nazornosti spet prikazan samo izračun za 50 tednov).



Grafikon 2: Sprememba tečaja USD (dejanski in napovedani iz modelov LeastMedSq in M5P)

Kot je razvidno iz grafikona 2, napovedni model narejen na podlagi algoritma M5P za malenkost bolje sledi dejanskim vrednostim sprememb menjalnega tečaja USD. To je razvidno tudi iz izračuna standardnega odklona razlike med dejanskimi in izračunanimi spremembami vrednostim tečaja, ki v tem primeru znaša 2,1076.

#### 4 IDEJE ZA IZBOLJŠAVO NAPOVEDNEGA MODELA

Iz dosedanjih izračunov im modeliranja v WEKI je razvidno, da ne obstaja enostavna povezava med spremembami vrednostmi tečaja USD in tečajem drugih valut. Torej spremembe tečajev ne moremo kar enostavno napovedati s spremembo tečajev ene ali dveh valut. Šele kombinacija več vrednosti tečajev valut nam da približen model za izračun tečaja USD. Iz tega sledi ideja, da verjetno niso vsi podatki o spremembah tujih valut enako informativni za določanje tečaja oziroma smeri gibanja tečaja USD. V ta namen imamo v WEKI pomoč, ki nam določi kateri atributi so najbolj informativni - »Select attributes« zavihek.

Kot cenilko atributov sem izbral cenilko »CfsSubsetEval«. Ta metoda nam določi "vrednost" posameznih atributov v podmnožici atributov, glede na njihovo napovedno sposobnost in korelacijo med njimi. Metoda, ki sem jo uporabil pri iskanju takih atributov pa je bila »genetsko iskanje« z 200 populacijami.

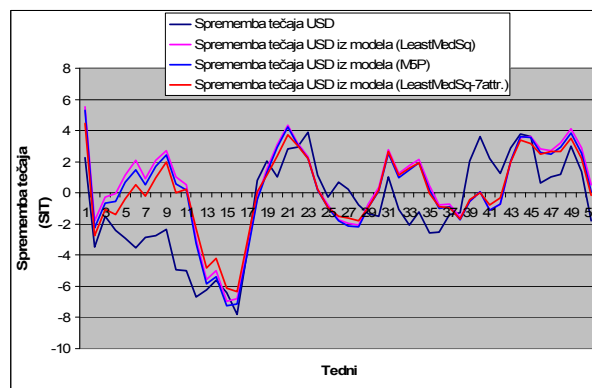
Rezultat takega iskanja so naslednje valute:

CAD, CZK, EUR, GBP, HRK, JPY, SKK

Torej to so najbolj informativne valute, ki nam pomagajo pri napovedi menjalnega tečaja USD. Sedaj lahko te attribute uporabimo pri izgradnji modela napovedi. Najprej sem uporabil linearno regresijo (LeastMedSq) in dobil naslednji model:

$$\text{USD} = [5] \begin{aligned} &0,695 * \text{CAD} + \\ &-6,7432 * \text{CZK} + \\ &-2,0119 * \text{EUR} + \\ &0,2572 * \text{GBP} + \\ &1,5974 * \text{HRK} + \\ &20,9834 * \text{JPY} + \\ &-1,2314 * \text{SKK} + \\ &0,0976 \end{aligned}$$

Praktični izračuna vrednosti sprememb menjalnega tečaja novega modela nam kaže grafikon 3 (zaradi nazornosti prikazan samo izračun za 50 tednov).



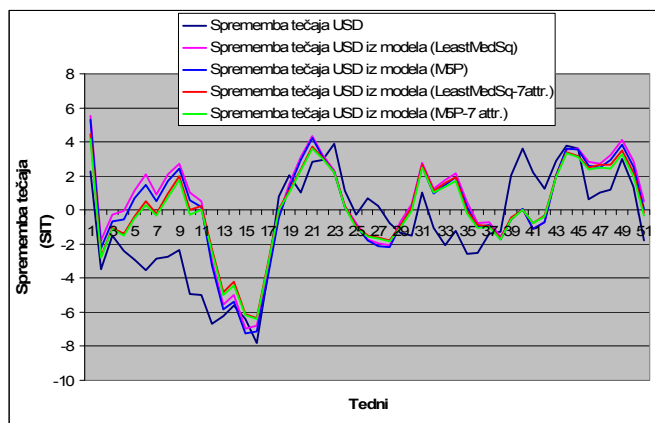
Grafikon 3: Sprememba tečaja USD (dejanski in napovedani iz modelov LeastMedSq (vseh in s 7 atributi) in M5P)

Kot lahko vidimo na grafikonu 3, smo z izbiro najbolj informativnih atributov še izboljšali model napovedi za spremembo menjalnega tečaja USD. Razlika je sicer relativno majhna, pa vendarle je, kar lahko vidimo tudi pri izračunu standardnega odklona razlike med dejanskimi in izračunanimi spremembami vrednostim tečaja, ki v tem primeru znaša 2,0325.

Če sedaj uporabimo isto logiko kot prej, da je načeloma M5P algoritem v našem primeru boljši (parameter *unpruned=false*), lahko izračunamo še model na podlagi tega algoritma. Rezultat je naslednji:

$$\text{USD} = [6] \begin{aligned} &0.7122 * \text{CAD} \\ &- 6.1727 * \text{CZK} \\ &- 1.5993 * \text{EUR} \\ &+ 0.242 * \text{GBP} \\ &+ 1.5397 * \text{HRK} \\ &+ 19.7 * \text{JPY} \\ &+ 0.1403 \end{aligned}$$

Če vrišemo v grafikon še rezultat tega modela, dobimo naslednjo sliko:



Grafikon 4: Sprememba tečaja USD (dejanski in napovedani iz modelov LeastMedSq (vseh in s 7 atributi) in M5P)

Iz grafikona 4 je razvidno, da smo model še za malenkost izboljšali, vendar zelo malo. Standardni odklon razlike med dejanskimi in izračunanimi spremembami vrednostim tečaja, sedaj znaša 1,9665.

To je tudi najboljši napovedni model, ki sem ga uspel izgraditi na izbranih podatkih.

## 5 IZGRADNJA PRAVEGA MODELNEGA DREVEESA

Seveda lahko na podlagi podatkov izgradimo tudi pravo modelno drevo. Tako drevo ima več listov (odvisno od parametra *minNumInstance*) in ima v vsakem listu svojo regresijo. Če parameter *minNumInstance* nastavimo na vrednost 100 (najmanj 100 primerov v vsakem listu) dobimo rezultat prikazan na sliki 1 spodaj.

Izkaže se, da je ta napovedni model manj točen kot na primer M5P algoritem (z rezanjem drevesa) s sedmimi atributi (npr. *Relative absolute error* je večji), tako, da je

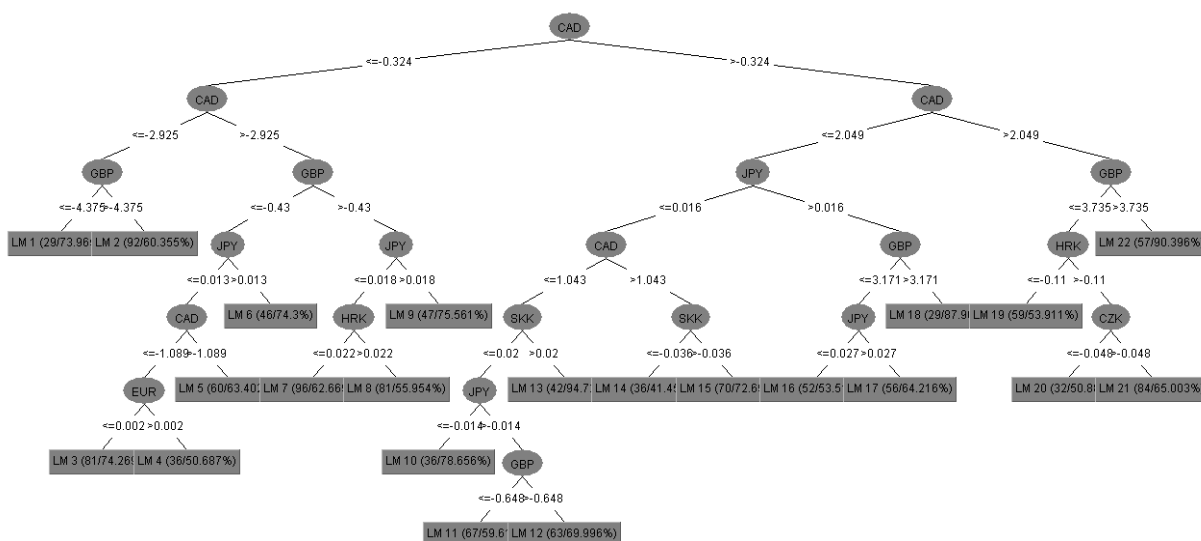
potrebno upoštevati v kakšnem primeru uporabljamo linearno regresijsko kdaj pa modelno drevo.

## 6 UGOTOVITVE IN ZAKLJUČKI

Cilj, te seminarske naloge je bil ugotoviti hipotetično povezavo med tečajem USD in ostalih 15 tujimi valutami. Ključna ugotovitev je, da enostavnega modela za napoved spremembe tečaja USD ni, ampak je potrebno vključiti veliko število vrednosti tujih valut. Konkretno to pomeni vrednosti tečajev sedmih tujih valut, ki so se pri določanju informativnosti izkazali kot najboljši (CAD, CZK, EUR, GBP, HRK, JPY, SKK). Izkaže se tudi, da je povezava med tečaji tujih valut prekompleksna, da bi si jo človek lahko razlagal na enostaven način. Iz modelov lahko vidimo, da imajo atributi različne vrednosti koeficientov v regresijski enačbi, kar pomeni, da imajo nekateri večji vpliv na rezultat kot drugi (npr. v vseh modelih bistveno izstopa japonski jen). Končni oziroma najboljši model, ki smo ga tako dobili s pomočjo »modelnega drevesa«, sicer deluje vendar gledano absolutno, kar s precejšnjo napako in bi se v praksi težko zanesli nanj. Model bi se še dal izboljšati s poznavanjem in uporabo zahtevnejših metod.

### Viri in literatura

- [1] Podatki o tečajih valut od leta 2002 do 2006. Pridobljeno s svetovnega spleta 10.04.2007 <http://www.bsi.si>
- [2] WEKA Explorer Guide. Pridobljeno s svetovnega spleta 05.04.2007 <http://www.cs.waikato.ac.nz/ml/weka/>
- [3] Na kratko o regresiji. Pridobljeno s svetovnega spleta 05.04.2007 <http://kt.ijs.si/petrakralj/UNGKnowledgeDiscovery0607.html>



Slika 1: Modelno drevo zgrajeno iz sedmih ključnih atributov