



Univerzitet u Novom Sadu
Prirodno - matematički fakultet,
Departman za fiziku

UPOTREBA PROGRAMSKI JEZIK PYTHON U ANALIZI METEOROLOŠKIH PODATAKA

-MASTER RAD-

Mentor: Prof.dr Ilija Arsenić

Kandidat: Martin Petraš

Novi Sad, 2018.

Sadržaj

1	Python i Django	2
2	Postupak za instaliranje	2
2.1	Provera verzije pythona	3
2.2	Izbegavanje problema sa dve verzije	4
3	Korišćenje biblioteka	5
3.1	NumPy	5
3.2	Pandas	6
3.3	Matplotlib	7
3.4	MetPy	7

Uvod

Predmet ovog rada je izrada web aplikacije u Django razvojnom okruženju za analizu meteoroloških podataka. Web aplikacija koristi podatke sa Carpatoclim projekta, projekat koji je prikupljao meteorološke podatke u vremenskom periodu od 1961-2010 godine. Obuhvaćena oblast, kao što i samo ime projekta kaže, je oblast pružanja planinskog lanca Karpatskih planina.

1 Python i Django

Programski jezik Python spada u interpreterske programske jezike visokog nivoa. Njegov autor je Gvido van Rosum. Njegova sintaksa omogućava pisanje veoma preglednih programa, brzo i lako učenje. Programski jezik Python nastao je krajem devedsetih godina prošlog veka. Broj funkcija u samom jeziku je skroman, pa zahteva relativno malo uloženog vremena i napora kako bi se napravio prvi programi. Pythonova sintaksa je dizajnirana da bude čitljiva i jednostavna, što ga čini idealnim nastavnim jezikom i omogućava početnicima brzo napredovanje. Programeri provode više vremena razmišljajući o problemu koji pokušavaju da reše, a manje vremena razmišljaju o kompleksnosti jezika. Najosnovnije korišćenje Pythona je kao jezik skriptovanja i automatizacije, koristi se za nauku o podacima i mašinsko učenje, za veb usluge, metaprogramiranje. Njegova popularnost se ogleda u velikom broju biblioteka. Što se tiče nedostataka istakli bi njegovu brzinu. Ono za što Pythonu treba šest sekundi u drugom programskom jeziku se može uraditi za delić sekunde. Međutim, brzina kojom se može napisati funkcionalan program je daleko brži nego u nekom drugom jeziku.

2 Postupak za instaliranje

Python kao programski jezik postoji za sve tri popularne systemske platforme Windows, Linux i OSX. Ovde će biti prikazan postupak instaliranja potrebnih paketa za rad pod Linux operativnim sistemom. Konkretno ćemo raditi na distribuciji baziranoj na Debian(stable) paketima, međutim za ostale distribucije poput Suse, Fedora, Arch postupak je sličan. Prvo što treba da uradimo jeste da odemo na <https://www.anaconda.com/download/#linux> i preuzmemo *Anaconda3-5.2.0-Linux-x86_64.sh*¹. U pitanju je *shell* skripta koja se pokreće u terminalu na sledeći način:

```
bash Anaconda3-5.2.0-Linux-x86_64.sh
```

Nakon što se prihvati licenca određujemo gde da instaliramo Anacondu, najbolje je da se ostavi podrazumevana putanja */home/USER/anaconda*. Nakon završene instalacije najbolje je uraditi modifikaciju *.bashrc* dodavanjem sledeće putanje:

```
export PATH=/home/$USER/anaconda/bin:$PATH
```

Nakon ovoga uspešno smo instalirali anacondu koju ćemo koristiti za instaliranje odnosno ažuriranje Python paketa. Paketi se uz pomoć Anaconde instaliraju

¹U pitanju je trenutna verzija

upisivanjem sledeće komande u terminal:

```
conda install ime_paketa
```

Za ažuriranje paketa koristi se sledeća komanda:

```
conda update ime_paketa
```

Postoji još jedan metod kojim se mogu instalirati paketi, a to je uz pomoć komande *pip*:

```
pip install ime_paketa
```

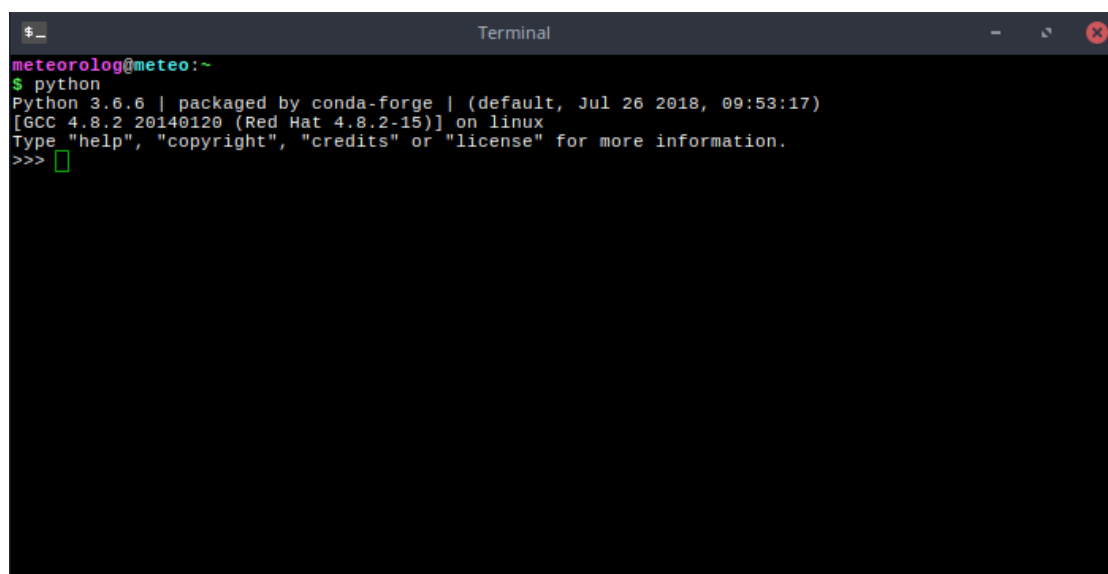
Međutim, jako bitno je istaći da ne treba mešati ova dva metoda za instaliranje paketa. Najbolje je da se izabere jedan od njih i uvek koristi, jer mogu nastati problemi ako se koriste oba metoda.

2.1 Provera verzije pythona

Prvo proveravamo da li imamo instaliran python kucanjem sledeće komande:

```
python
```

Ukoliko je python instaliran pojavi se sledeće Uglavnom sve distribucije linuxa



```
Terminal
meteorolog@meteo:~
$ python
Python 3.6.6 | packaged by conda-forge | (default, Jul 26 2018, 09:53:17)
[GCC 4.8.2 20140120 (Red Hat 4.8.2-15)] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> █
```

Slika 1. Pokretanje pythona

dolaze sa instaliranim pythonom, razlika može biti u verziji. Postoje dve verzije pythona, 2.7 čija podrška polako prestaje i novija 3.x (trenutna je 3.6.6). Razlike između ove dve verzije postoje, ali metpy biblioteka koja nama je potrebna podržava obe verzije. Više o razlikama između ove dve verzije pogledajte

na <https://wiki.python.org/moin/Python2orPython3>. Koja verzija pythona je instalirana možemo proveriti i kucanjem sledećeg koda u terminal:

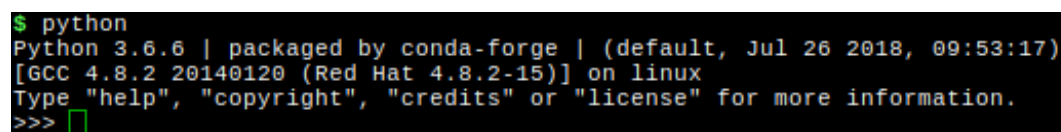
```
python --version  
Python 3.6.6
```

U buduće svaki kod će biti vezan za ovu verziju.

2.2 Izbegavanje problema sa dve verzije

Kao što sam već spomenuo sve linux distribucija dolaze sa instaliranim pythonom, razlike mogu biti u verziji. Ukoliko vaša linux distribucija ima instaliranu samo verziju 2.7 a vi bi hteli i verziju 3.x, onda treba obratiti pažnju na sledeće. Kod mene, ja koristim MX linux distribuciju baziranu na Debian(stable) verziji, kucanjem u terminalu:

```
python --version  
Python 3.6.6
```

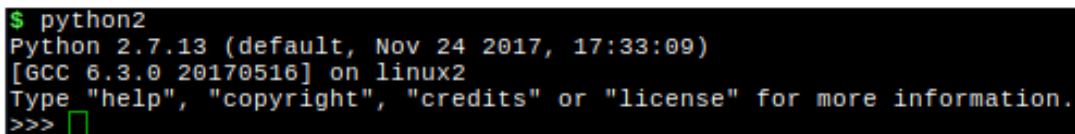


```
$ python  
Python 3.6.6 | packaged by conda-forge | (default, Jul 26 2018, 09:53:17)  
[GCC 4.8.2 20140120 (Red Hat 4.8.2-15)] on linux  
Type "help", "copyright", "credits" or "license" for more information.  
>>> 
```

Slika 2. Pokretanje pythona verzije 3.6.6

prikazuje da je podrazumevana verzija 3.6. Međutim, ukoliko je kod vas instalirana samo verzija 2.7 a vi dodatno instalirate verziju 3.x, može se desiti da će verzija 2.7 i dalje ostati podrazumevana, pa se prilikom kompajliranja može javiti problem. Pošto su kod mene obe verzije, kao podrazumevana verzija je 3.6.6 dok je za pozivanje verzije 2.7 potrebno kucati sledeće:

```
python2 --version  
Python 2.7.13
```



```
$ python2  
Python 2.7.13 (default, Nov 24 2017, 17:33:09)  
[GCC 6.3.0 20170516] on linux2  
Type "help", "copyright", "credits" or "license" for more information.  
>>> 
```

Slika 3. Pokretanje pythona verzije 2.7

Kako bi otklonili ovaj problem, u direktorijumu `/home/user/.bashrc`, koristeći jedna od tekst editora, upišemo sledeću komandu:

```
alias python=python3
```

Ovom komandom prilikom svakog pozivanja pythona iz terminala kao podrazumevana verzija će biti 3.6.6, što će biti jako bitno u daljem radu.

3 Korišćenje biblioteka

Mnoge python funkcije su sadržane u specijalizovanim bibliotekama, tzv. modulima. Učitavanje modula se postiže naredbom *import*. Ova mogućnost je pythonu donela popularnost jer postoje mnogo modula koji jako pomažu prilikom rada. Python je svojom popularnošću stekao pri izradi veb programa međutim usavršavanjem podrške za dodatne module otvorilo je vrata pythonu i u drugim oblastima. Moduli poput *numpy* i *pandas*, koji se koriste za analizu i vizuelno prikazivanje podataka, su python svrstali uz rame sa ostalim kako komercijalnim programima tako i sa programima otvorenog koda kao što su R, MATLAB, SAS, Stata i ostali. Kada pogledamo da python spada u besplatni programski jezik, njegova popularnost nije slučajnost. Jedan deo ove popularnosti u obradi podataka pripada i laka integracija sa C, C++ i FORTRAN kodom. Pošto su moduli jako značajni za python istaći ćemo neke od najznačajnijih. Samo da napomene, sve nabrojane mogućnosti koje pružaju biblioteke je moguće izvesti koristeći isključivo samo Pythonov kod. Prednost biblioteka je u efikasnosti i lakoći kojim se postižu isti rezultati, uz manje utrošenog vremena.

3.1 NumPy

NumPy (*Numerical Python*) predstavlja fundamentalnu open source Python biblioteku kada su obzir uzmu numerički proračuni. Sadrži matematičke funkcije zadužene za operaciju na podacima, koje se veoma brzo i efikasno izvršavaju. Poziva se komandom :

```
import numpy as np
```

gde *np* predstavlja skraćenicu od *numpy*. Kao primer koristićemo funkciju *np.linspace* kojom se generišu brojevi u zadatom intervalu.

Za numeričke proračune, nizovi kod NumPy su dosta efikasniji prilikom njihove manipulacije nego bilo koja druga izgrađena struktura unutar Pythona. Instaliranje *numpy* biblioteke se radi na sledeći način:

```
conda install numpy
```

Ažuriranje novije verzije se radi na sledeći način:

```
conda update numpy
```

```
In [1]: import numpy as np
In [2]: np.linspace(0,2,9)
Out[2]: array([0. , 0.25, 0.5 , 0.75, 1. , 1.25, 1.5 , 1.75, 2. ])
```

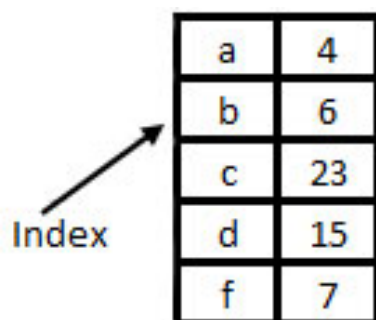
Slika 4. Pozivanje numpy funkcije `np.linspace` za generisanje devet brojeva u opsegu od nula do dva.

3.2 Pandas

Pandas je dizajniran kako bi se ubrzao rad sa strukturiranim ili tabelarnim podacima i kao takav, postao jako moćna i produktivna alatka za analizu podataka. Ono što krasi ovu biblioteku je i velika podrška od strane zajednice, njen aktivna razvoj, odličan rad sa ostalim bibliotekama, izgrađena je na osnovama *numpy* pa je takođe krasi brzina. Neke od mogućnosti koje pruža ova biblioteka su:

- ulaz-izlaz podataka u različitim formatima (csv, txt, SQL...)
- indeksiranje, sortiranje, rangiranje
- čišćenje podataka
- grupisanje
- vizuelizacija

Sačinjenja je od dve strukture podataka, *dataframe* i *series*. *Series* predstavljaju jednodimenzioni objekat sačinjen od tabele sa vrednostima i njihovim indeksima. *Dataframe* predstavlja tabelarnu strukturu podataka, definisanih u dve



a	4
b	6
c	23
d	15
f	7

Slika 5. Pandas series

ili više dimenzije. Sačinjen je od tri osnovne komponente: podataka, indeksa, kolona. Skraćenica za *pandas* je *pd*, poziva se korišćenjem komande:

	Žanr	Ime pesme	Godina izdanja
0	Disco	The Trammps - Disco Inferno	1976
1	Soul	Patti LaBelle - Funky Music	1977
2	Funk	Ohio Players - Love Rollercoaster	1975

Slika 6. Pandas dataframe

```
import pandas as pd
```

Kao primer uzeću mogućnost ubacivanja podataka iz .csv dokumenta, to se radi na sledeći način:

```
df = pd.csv_read('ime_dokumenta.csv')
```

Instaliranje pandas biblioteke:

```
conda install pandas
```

3.3 Matplotlib

Najviše korišćena i najstarija biblioteka koja se koristi za vizuelno predstavljanje podataka. U kombinaciji sa *numpy* i *pandas* čine jako moćne alate, koji se mogu nositi sa komercijalnim i dosta skupim alatima kao što su *Matlab* i *Mathematica*. Mana ove biblioteke je potreba za pisanjem većeg broja linija koda kako bi se dobila naprednija vizuelizacija podataka. Potreba za povećanim kvalitetom vizuelizacije, nastale su savremenije biblioteke koje se u većoj ili manjoj meri oslanjaju na *matplotlib*. Jedna od takvih biblioteka je *seaborn* čiji je fokus na atraktivnijoj izradi statističkih grafika. *Seaborn* se potpuno oslanja na *matplotlib*. Naveo bih još i *bokeh* i *altair* koje nisu zavisne od *matplotlib*, a takođe se koriste za vizuelizaciju podataka.

3.4 MetPy

Metpy predstavlja kolekciju alata u Pythonu za pregled, vizuilizaciju i proračuna meteoroloških podataka. Paket je nastao od strane *Unidata* udruženja, koja se bavi izradom alata za naučne svrhe. MetPy paket sadrži dosta opcija, koristi se za vizelizaciju radarskih slika, rad sa netCDF podacima, vizuelizaciju radio-sondažnih merenja, vizuelizaciju meteoroloških podataka i ostalo. Mi ćemo koristiti ovu biblioteku u svrhu interpolacije polja temperature. Za instaliranje MetPy u terminal kucamo:

```
conda install -c -conda-forge metpy
```

Instaliranjem ovog paketa dobijamo razne alate, nama će najviše značiti mogućnost interpolacije za polje temperature kao i njegove vizuelizacije. Za vizuilizaciju se koristi *Cartopy*, Pythonov paket za crtanje mapa.