

Uvod u Jupyter Notebook i Python za strojno učenje

Bruno Polonijo, asistent

Što je Jupyter Notebook?

- Interaktivno računalno okruženje.
- Omogućuje kreiranje i dijeljenje dokumenata koji sadrže živi kod, jednačbe, vizualizacije i tekst.
- Popularan među znanstvenicima podataka, istraživačima, obrazovnim ustanovama i mnogim drugima

Prednosti korištenja Jupyter Notebooka

- Interaktivnost: izvršavanje koda u stvarnom vremenu.
- Podrška za više jezika programiranja (npr. Python, R, Julia).
- Lako dijeljenje: prezentacija rezultata, izrada izvješća ili dijeljenje s drugima putem interneta.
- Vizualizacija podataka: integracija s popularnim alatima poput Matplotlib, Seaborn itd.

Python i strojno učenje

- Python je jedan od najpopularnijih jezika za strojno učenje.
- Bogat ekosustav biblioteka kao što su Pandas, Scikit-Learn, TensorFlow, Keras i druge.
- Aktivna zajednica koja pruža podršku, tutorijale i resurse.

Osnove Pythona za strojno učenje

- Učitavanje podataka: pandas
- Obrada podataka: numpy i pandas
- Vizualizacija podataka: matplotlib i seaborn
- Modeliranje: Scikit-learn

Početak rada s Jupyter Notebookom (JupyterLab)

- **JupyterLab:** Napredna verzija tradicionalnog Jupyter Notebook-a s boljim korisničkim sučeljem i dodatnim funkcionalnostima.
- **Kako instalirati:**
 - Pomoću pip: “pip install jupyterlab”
 - Ako koristite Anaconda: “conda install -c conda-forge jupyterlab”

Pokretanje JupyterLab-a iz AppData foldera

1. **Otvorite Command Prompt (CMD).**
2. **Navigirajte do foldera gdje je Python instaliran, obično:**
C:\Users\[Vaše korisničko ime]\AppData\Local\Programs\Python\[Verzija Pythona]
3. **Pokrenite JupyterLab unosom: jupyter-lab**

Alternativa - Korištenje Anaconda

- **Anaconda:** Popularna distribucija za znanstvene svrhe koja uključuje Python, Jupyter i mnoge druge pakete.
- **Prednosti korištenja Anaconda:**
 - Lako upravljanje paketima i okruženjima.
 - Ugrađen JupyterLab i mnoge druge znanstvene biblioteke.
- **Kako koristiti:**
 - Preuzmite i instalirajte Anaconda s njihove službene web [stranice](#).
 - Nakon instalacije, otvorite **Anaconda Navigator** i pokrenite JupyterLab.

Osnovne značajke JupyterLab-a

- **Više dokumenata:** Možete otvoriti više bilježnica, tekstualnih datoteka i konzola u istoj sesiji.
- **Proširivost:** Mogućnost dodavanja dodataka kako biste proširili funkcionalnosti.
- **Interaktivne ekstenzije:** Podrška za interaktivne grafičke sučelja, poput widgeta.
- **Integracija s drugim alatima:** Lako se povezuje s Gitom, može se koristiti za uređivanje Markdown datoteka, JSON i još mnogo toga.

Primjer korištenja JupyterLaba 1.

- Kreirajte u excelu csv datoteku s podacima i nazovite datoteku test.csv:

Prodavnica	Proizvod	Kolicina prodana	Cijena	Promocija
A	Jabuka	100	0.5	Da
A	Banana	150	0.3	Ne
B	Jabuka	80	0.55	Ne
B	Banana	90	0.35	Ne

- Upalite JupyterLab i unesite iduće komande da biste učitali podatke

```
import pandas as pd #služi uključivanju programske biblioteke koja služi obradi i analizi podataka
data = pd.read_csv('C://Users/abhothoh/Documents/test.csv') # napravite novu varijablu (data) i dodijelite podatke iz test.csv
```


Primjer korištenja JupyterLaba 2.

- Prikaz podataka:

```
print(data) #prikazuje podatke spremljene u varijabli
```

	Prodavnica	Proizvod	Kolicina prodana	Cijena	Promocija
0	A	Jabuka	100	0.50	Da
1	A	Banana	150	0.30	Ne
2	B	Jabuka	80	0.55	Ne
3	B	Banana	90	0.35	Ne

- Prikaz informacija o spremljenim podacima

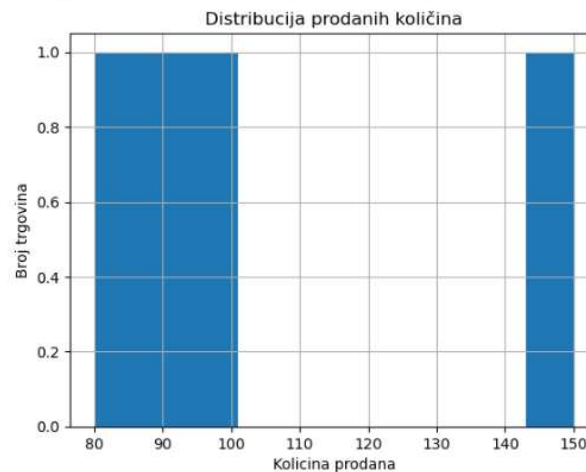
```
print(data.info()) #Prikazuje informacije o spremljenim podacima (imena stupaca, vrsta podatka i slično).
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4 entries, 0 to 3
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Prodavnica      4 non-null     object
1   Proizvod        4 non-null     object
2   Kolicina prodana 4 non-null     int64
3   Cijena          4 non-null     float64
4   Promocija       4 non-null     object
dtypes: float64(1), int64(1), object(3)
memory usage: 292.0+ bytes
None
```

Primjer korištenja JupyterLaba 3.

- Vizualizacija podataka:

```
import matplotlib.pyplot as plt #služi uključivanju programske biblioteke za vizualizaciju podataka  
data['Kolicina prodana'].hist() #prezicira podatak koji će se prikazat u grafikonu  
plt.title('Distribucija prodanih količina') #Dodaje naziv grafikonu  
plt.xlabel('Kolicina prodana') #Dodaje naziv x osi  
plt.ylabel('Broj trgovina') #Dodaje naziv y osi  
plt.show() #Služi prikazu konfiguriranog grafikona
```



Primjer korištenja JupyterLaba 3.

- Prikaz prosjeka promocija (da/ne):

```
print(data.groupby("Promocija")["Cijena"].mean()) #provjera prosječne cijene proizvoda kada je na promociji nasuprot kada nije:
```

Promocija
Da 0.5
Ne 0.4
Name: Cijena, dtype: float64

- Dodavanje novog stupca koji prikazuje ukupnu prodaju:

```
data["Ukupna prodaja"] = data["Kolicina prodana"] * data["Cijena"]  
print(data) #dodavanje novog stupca koji prikazuje ukupnu prodaju
```

	Prodevnica	Proizvod	Kolicina prodana	Cijena	Promocija	Ukupna prodaja
0	A	Jabuka	100	0.50	Da	50.0
1	A	Banana	150	0.30	Ne	45.0
2	B	Jabuka	80	0.55	Ne	44.0
3	B	Banana	90	0.35	Ne	31.5

Primjer korištenja JupyterLaba 4.

- Kreiranje kolone ukupna prodaja i sumiranje po proizvodu:

```
data['Ukupna prodaja'] = data['Kolicina prodana'] * data['Cijena']  
sum_by_product = data.groupby('Proizvod')['Ukupna prodaja'].sum()  
print(sum_by_product)
```

Proizvod

Banana 76.5

Jabuka 94.0

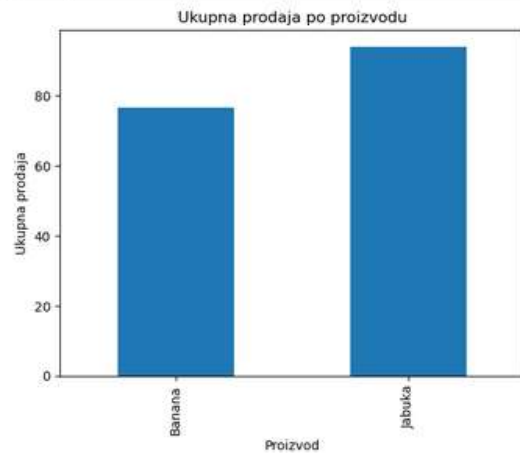
Name: Ukupna prodaja, dtype: float64

Primjer korištenja JupyterLaba 5.

- Prikaz ukupne prodaje po proizvodu:

```
import matplotlib.pyplot as plt

sum_by_product.plot(kind='bar') #kao i prije specificiramo podatak koji se koristi za grafikon, ali dodatno u (kind=bar) specificiramo koju vrstu grafikona zelimo
plt.title('Ukupna prodaja po proizvodu')
plt.ylabel('Ukupna prodaja')
plt.show()
```



Zaključak

- Jupyter Notebook je moćan alat za analizu podataka i strojno učenje.
- Python pruža sve potrebne alate i biblioteke za efikasno i uspješno bavljenje strojnim učenjem.
- Početak je jednostavan, ali mogućnosti za učenje i rast su beskrajne!

Pitanja?