Adding_TargetPrefName_To_Predictions

February 14, 2024

1 Adding Target Pref Name Column To Our Predictions

1.1 Imports

[2]:
```
import os
import pandas as pd

def PRINT(text) -> None: print(f"{80*'-'}\n{text}\n{80*'-'}")
```

1.2 Load the Datasets

[3]:
```
timbal_dataset_df = pd.read_csv(os.path.join('data', 'timbal_triplets.csv'))

PRINT(f'Loaded Timbal dataset csv file to pandas data frame successfully')
```

```
--------------------------------------------------------------------------------
Loaded Timbal dataset csv file to pandas data frame successfully
--------------------------------------------------------------------------------
```

[4]:
```
timbal_dataset_df.head()
```

[4]:

	timbal_v2_id	smiles	target_name	uniprot_target
0	11821	CC(O)CN1C(C(=C(O)C1=O)C(=O)c2ccc(C)cc2)c3ccc(c…	Annexin A2	P60903
1	11864	COCCCN1C(C(=C(O)C1=O)C(=O)c2ccc(C)nc2)c3ccc(cc…	Annexin A2	P60903
2	16986	CC(O)CN1C(C(=C(O)C1=O)C(=O)c2ccc(C)cc2)c3ccc4c…	Annexin A2	P07355
3	16973	CCOc1ccccc1C(=O)C2=C(O)C(=O)N(CCCOC)C2c3ccc(cc…	Annexin A2	P07355
4	11861	COCCCN1C(C(=C(O)C1=O)C(=O)c2ccc(C)nc2)c3ccc(c(…	Annexin A2	P60903

[14]:
```
temp_timbal = timbal_dataset_df[['smiles', 'target_name', 'uniprot_target']]
```

[22]:
```
temp_timbal.drop_duplicates(inplace=True)
```

```
C:\Users\gavvi\AppData\Local\Temp\ipykernel_28788\3313701805.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  temp_timbal.drop_duplicates(inplace=True)
```

[23]:
```
PRINT(f'Timbal data frame shape after dropping -> {temp_timbal.shape}')
```

```
--------------------------------------------------------------------------------
Timbal data frame shape after dropping -> (8703, 3)
--------------------------------------------------------------------------------
```

[24]:
```
predictions_df = pd.read_csv('predictions.csv')

PRINT(f'Loaded predictions csv file into pandas data frame successfully !')
```

```
--------------------------------------------------------------------------------
Loaded predictions csv file into pandas data frame successfully !
--------------------------------------------------------------------------------
```

[25]:
```
predictions_df.head()
```

[25]:

	SMILES	UniProtTarget	PredictedUniProtPartner
0	OC(=O)[C@H](Cc1ccc(NC(=O)c2c(Cl)cccc2Cl)cc1)NC…	P13612	P05556
1	CC1CCC(C[C@H](NC(=O)[C@@H]2CCC(=O)N2Cc3ccccc3)…	P13612	P05556
2	CC(C)CCNC(=O)[C@@H]1OCO[C@H]1C(=O)N[C@@H](Cc2c…	P13612	P05556
3	OC(=O)CN(CC(=O)N[C@@H](Cc1ccc(OCc2c(Cl)cccc2Cl…	P13612	P05556
4	CCC\N=C/1\C(\C(=C1O)O)=N\[C@@H](Cc2ccc(OCc3c(C…	P13612	P05556

1.3 Merge the Data Frames In Order to Extract Target Pref Name

[26]:
```
timbal_columns = list(temp_timbal.columns)
predictions_columns = list(predictions_df.columns)
```

[27]:
```
PRINT(f'Timbal data frame columns -> {timbal_columns}\n\nPredictions data frame␣
  ↪columns -> {predictions_columns}')
```

```
--------------------------------------------------------------------------------
Timbal data frame columns -> ['smiles', 'target_name', 'uniprot_target']
```

```
Predictions data frame columns -> ['SMILES', 'UniProtTarget',
'PredictedUniProtPartner']
--------------------------------------------------------------------------------
```

[28]:
```
merged_df = pd.merge(predictions_df, temp_timbal, left_on=['SMILES',␣
↪'UniProtTarget'], right_on=['smiles', 'uniprot_target'], how='left')

PRINT(f'Merged the data frames by molecule SMILES value successfully !')
```

```
--------------------------------------------------------------------------------
Merged the data frames by molecule SMILES value successfully !
--------------------------------------------------------------------------------
```

[34]:
```
PRINT(f'Verify that number of samples of predictions data frame and our merged␣
↪data frame equal:\n\nPredictions -> {predictions_df.shape[0]}, Merged ->␣
↪{merged_df.shape[0]}')
```

```
--------------------------------------------------------------------------------
Verify that number of samples of predictions data frame and our merged data
frame equal:

Predictions -> 4192, Merged -> 4192
--------------------------------------------------------------------------------
```

[35]:
```
merged_df.rename(columns={'target_name': 'Target Pref Name'}, inplace=True)
```

[36]:
```
merged_df.head()
```

[36]:

	SMILES	UniProtTarget	\
0	OC(=O)[C@H](Cc1ccc(NC(=O)c2c(Cl)cccc2Cl)cc1)NC…	P13612	
1	CC1CCC(C[C@H](NC(=O)[C@@H]2CCC(=O)N2Cc3ccccc3)…	P13612	
2	CC(C)CCNC(=O)[C@@H]1OCO[C@H]1C(=O)N[C@@H](Cc2c…	P13612	
3	OC(=O)CN(CC(=O)N[C@@H](Cc1ccc(OCc2c(Cl)cccc2Cl…	P13612	
4	CCC\N=C/1\C(\C(=C1O)O)=N\[C@@H](Cc2ccc(OCc3c(C…	P13612	

	PredictedUniProtPartner	smiles	\
0	P05556	OC(=O)[C@H](Cc1ccc(NC(=O)c2c(Cl)cccc2Cl)cc1)NC…	
1	P05556	CC1CCC(C[C@H](NC(=O)[C@@H]2CCC(=O)N2Cc3ccccc3)…	
2	P05556	CC(C)CCNC(=O)[C@@H]1OCO[C@H]1C(=O)N[C@@H](Cc2c…	
3	P05556	OC(=O)CN(CC(=O)N[C@@H](Cc1ccc(OCc2c(Cl)cccc2Cl…	
4	P05556	CCC\N=C/1\C(\C(=C1O)O)=N\[C@@H](Cc2ccc(OCc3c(C…	

	Target Pref Name	uniprot_target
0	Integrins	P13612
1	Integrins	P13612
2	Integrins	P13612
3	Integrins	P13612
4	Integrins	P13612

[38]:
```
PRINT(f'Merged data frame columns are -> {list(merged_df.columns)}')
```

```
--------------------------------------------------------------------------------
Merged data frame columns are -> ['SMILES', 'UniProtTarget',
'PredictedUniProtPartner', 'smiles', 'Target Pref Name', 'uniprot_target']
--------------------------------------------------------------------------------
```

[39]:
```
res_df = merged_df[['SMILES', 'UniProtTarget', 'PredictedUniProtPartner',␣
↪'Target Pref Name']]
```

[41]:
```
res_df.head(2)
```

[41]:

	SMILES	UniProtTarget	PredictedUniProtPartner	Target Pref Name
0	OC(=O)[C@H](Cc1ccc(NC(=O)c2c(Cl)cccc2Cl)cc1)NC…	P13612	P05556	Integrins
1	CC1CCC(C[C@H](NC(=O)[C@@H]2CCC(=O)N2Cc3ccccc3)…	P13612	P05556	Integrins

1.4 Save

[42]:
```
res_df.to_csv('predictions_with_tpf.csv', index=False)

PRINT('Saved !')
```

```
--------------------------------------------------------------------------------
Saved !
--------------------------------------------------------------------------------
```

[ ]:
```

```