

TEAM ID: PNT2022TMID31349

Resource list

Name	Group	Location	Product	Status	Tags
Filter by name or IP address... Filter by group or org... Filter... Filter... Filter... Filter...					
Compute (0)					
Containers (0)					
Networking (0)					
Storage (1)					
Cloud Object Storage-mr	Default	Global	Cloud Object Storage	Active	cpdaas
AI / Machine Learning (2)					
Watson Machine Learning-ce	Default	Dallas	Watson Machine Learning	Active	-
Watson Studio-qv	Default	Dallas	Watson Studio	Active	-
Analytics (0)					
Blockchain (0)					
Databases (0)					
Developer tools (0)					

University Prediction

```
In [35]: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=1)

MULTIPLE LINEAR REGRESSION

In [36]: mul_lin_reg = LinearRegression()
mul_lin_reg.fit(x_train, y_train)
y_pred_mlr = mul_lin_reg.predict(x_test)
r2_score_mlr = r2_score(y_test, y_pred_mlr)
print("Multiple Linear Regression's Score = {:.3f}".format(r2_score_mlr))

Multiple Linear Regression's Score = 0.888

RANDOM FOREST REGRESSION

In [37]: ran_for_reg = RandomForestRegressor(n_estimators=100, random_state=1)
ran_for_reg.fit(x_train, y_train)
y_pred_rfr = ran_for_reg.predict(x_test)
r2_score_rfr = r2_score(y_test, y_pred_rfr)
print("Random Forest Regression's Score = {:.3f}".format(r2_score_rfr))

Random Forest Regression's Score = 0.778

In [38]: import pickle
pickle.dump(mul_lin_reg, open("MULTIPLE_LINEAR_REGRESSION.pkl", "wb"))

Conclusion

R^2 score is an indicator of accuracy of Regression Models, and the accuracy is measured as close to 1 of this value. Therefore, as seen, Multiple Linear Regression Model is better than Random Forest Regression on this dataset when comparing their R^2 scores.

SAVE THE MODEL IN IBM CLOUD

In [39]: !pip install ibm_watson_machine_learning

Requirement already satisfied: ibm_watson_machine_learning in /opt/conda/envs/Bython3.9/lib/python3.9/site-packages (1.0.107)
```