أكاديمية سدايا
SDAIA Academy

Project report : Polynomial regression model
Predict universities scores

By
Bushra AlOliyt
Laila Almjnuni

Data Science Bootcamp SDAIA Academy

Abstract

The goal is to improve the Saudi universities ranking by knowing which feature has a greater impact as universities is the most important place to have knowleage and learning in the best way , we start web-scraped on the Top Universities to collect the whole universties over the world and identify with feature has the greater effect on the score point to know how to impact our saudi universities by applying this features and studyng it

Design

Apartment owners face difficulty in determining the right effect on the score of universities , with score competition between apartment owners, apartment owners need a way to help them determine the features that has the greater impact on score the consumer, based on the specific characteristics of each property

Algorithms

1. Read ataframe from pickle after web-scraping
2. Cleaning data (removing missing values , unnecassery data)
3. Settting index
4. Apply diffrent model to know the best fit
5. predict y_train through unseendata X_test

Tools

- Numpy and Pandas for data manipulation
- Matplotlib and Seaborn for plotting
- Selenuim
- Sklearn

Data Source

We have used Web Scrapping to collect our data and the source was Top Universities

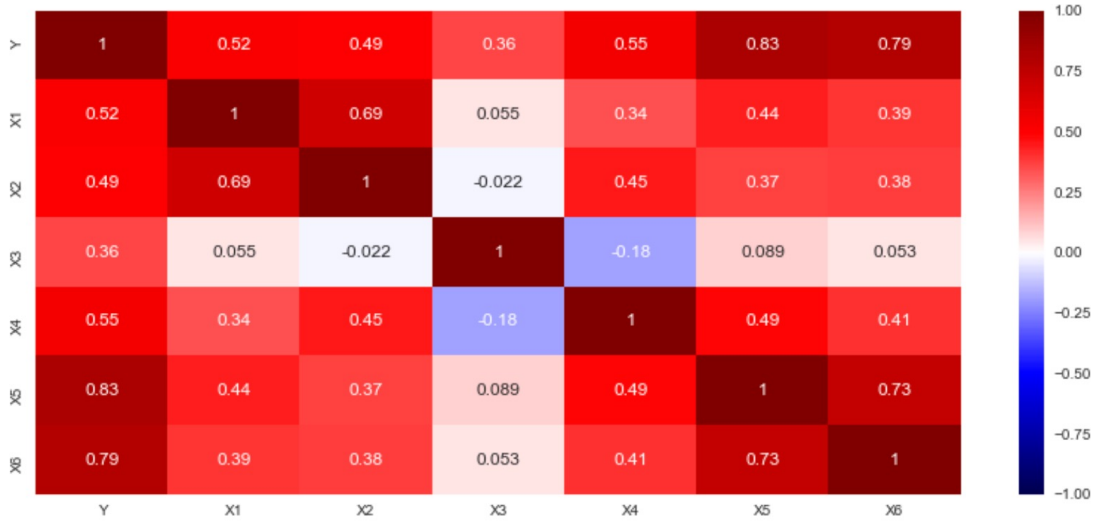
Communication

In addition to the slides presented,these are plots and visualizations

○ Data columns

Columns	Type
University Name	**String**
International Student Ratio	**Float**
International Faculty Ratio	**Float**
Faculty Student Ratio	**Float**
Citation Per Faculty	**Float**
Academic Reputation	**Float**
Employer Reputation	**Float**

- **Relations**



- **Results**

Models ussing Cross-Validation	X1,X2,X3,X4,X5,X6	R^2
Linear Regression	All features	Test R^2 = 0.88
Polynomial Regression	All features	Test R^2 = 0.89
Ridge Regression	All features	Test R^2 = 0.86
Lasso Regression	All features	Test R^2 = 0.86
Elastic Net Regression	All features	Test R^2 = 0.86
Random Forest Regression	All features	Test R^2 = 0.88

Using Coeffeitint formula for prediction

$$\hat{y} = \beta_0 + \beta_{X1} + \beta_{X2} + \beta_{X3} + \beta_{X4} + \beta_{X5} + \beta_{X6}$$

Where
- X1 → International Student Ratio
- X2 → International Faculty Ratio
- X3 → Faculty Student Ratio
- X4 → Citation Per Faculty
- X5 → Academic Reputation
- X6 → Employer Reputation

- **Prediction**

	Real Score	PolynomialRegression
0	500	20
1	200	18
2	300	20
3	400	19
4	100	19

	Real Score	PolynomialRegression
0	90	20
1	80	18
2	70	20
3	60	19
4	50	19