UNIVERSITY RANKING 2023

Presented By : Sarah Ali

INTRODUCTION

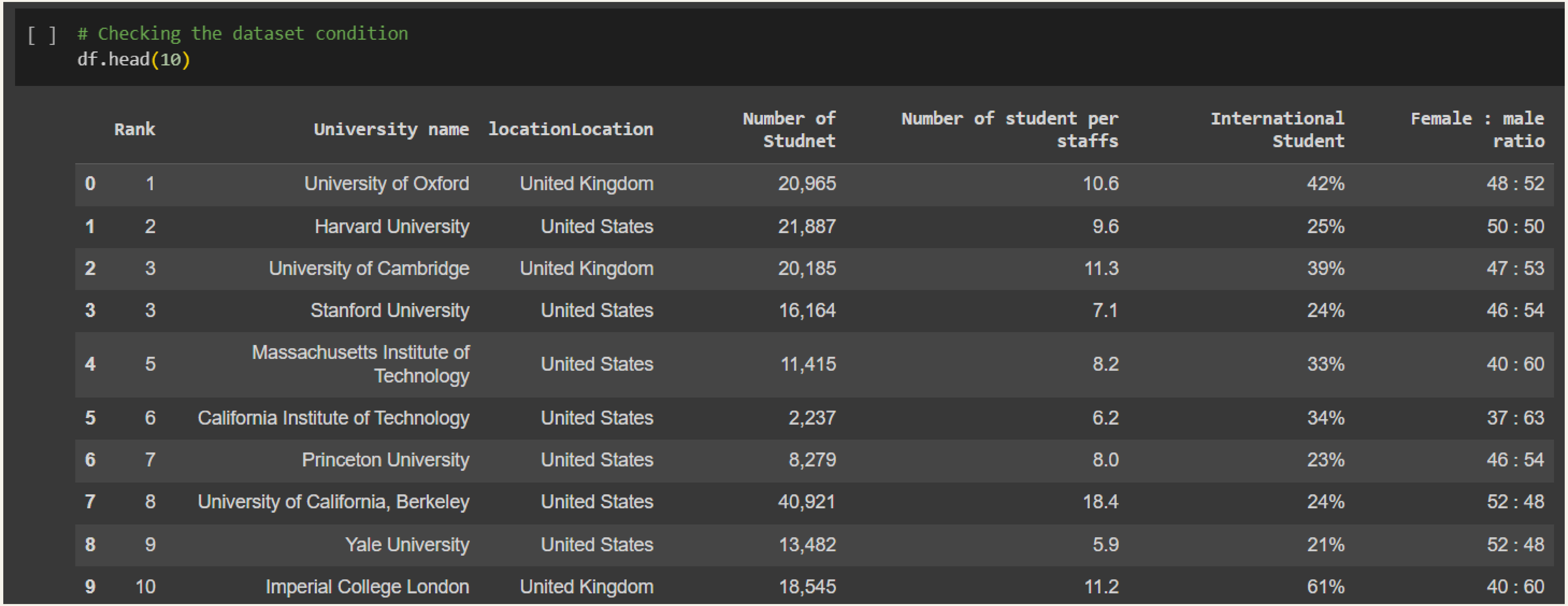
To accurately predict the likelihood of 'Number of Studnet' based on key features such as Rank, University name, Location, Number of Students, Number of Students per Staff, International Student Proportion, and Female to Male Ratio.
To know if there is a significant association between features or not.

DATASET DISCRIPTION

The "Global University Rankings Dataset 2023" is a comprehensive collection of key metrics and characteristics for top universities worldwide. The dataset provides insights into the performance and demographics of renowned academic institutions on a global scale.

- **Included features:**

[Rank, University name, Location, Number of Students,
Number of Students per Staff, International Student Proportion,
Female to Male Ratio]



```
[ ] # Checking the dataset condition
    df.head(10)
```

	Rank	University name	locationLocation	Number of Studnet	Number of student per staffs	International Student	Female : male ratio
0	1	University of Oxford	United Kingdom	20,965	10.6	42%	48 : 52
1	2	Harvard University	United States	21,887	9.6	25%	50 : 50
2	3	University of Cambridge	United Kingdom	20,185	11.3	39%	47 : 53
3	3	Stanford University	United States	16,164	7.1	24%	46 : 54
4	5	Massachusetts Institute of Technology	United States	11,415	8.2	33%	40 : 60
5	6	California Institute of Technology	United States	2,237	6.2	34%	37 : 63
6	7	Princeton University	United States	8,279	8.0	23%	46 : 54
7	8	University of California, Berkeley	United States	40,921	18.4	24%	52 : 48
8	9	Yale University	United States	13,482	5.9	21%	52 : 48
9	10	Imperial College London	United Kingdom	18,545	11.2	61%	40 : 60

METHODOLOGY

Preprocessing phase

- Data cleaning
  - missing value treatment
  - Editing columns names
  - Editing columns types

Data analysis phase

- Getting min, max, var, std, skewness, kurtosis

Visualization phase

- Numeric values distribution
- Categorical values statistics

Preprocessing phase

- Covariance matrix
- Correlation matrix
- Z-test/T-test
- ANOVA
- Chi-square

METHODOLOGY

Feature reduction phase

- LDA
  - With accuracy 0.99
- PCA
  - With accuracy 0.67
- SVD
  - With accuracy 0.04
- K-fold validation
  - Average aaccuracy 1.0

MODELS IMPLEMENTED

- **Naive Bayes**

  The accuracy is approximately 57%, meaning that the model correctly predicted the location category for 57% of the instances in the test set.

- **Bayesian Belief Network**

  Definning a Bayesian Network with nodes. The CPDs provide the conditional probabilities of each variable given its parent variable.

MODELS IMPLEMENTED

● Decision Tree

With accuracy 0.37

- The model correctly predicted the target variable about 37% of the time.

● Neural Networks

Last neural network result is:

```
Squared Error on test set:
1043149760.0 2/2
[=====================
=========] - 0s 11ms/step
```

MODELS IMPLEMENTED

- **K-Nearest Neighbors**

  With accuracy 0.03

  - **The model's performance in predicting the 'female_prop' target variable.**

- **K-fold cross validation and average accuracy**

  Accuracy Scores for Each Fold: [1. 1.] Average Accuracy: 1.0

  - **resulting in perfect accuracy across all folds.**

CONCLUSION

There is a significant association between rank and international students and also between location and international students!

REFERENCES

- Saritas, M. M., & Yasar, A. (2023, March 6). Performance analysis of ann and naive Bayes classification algorithm for Data Classification. International Journal of Intelligent Systems and Applications in Engineering. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque non elit mauris. Cras euismod, metus ac finibus.
- Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (1970, January 1). KNN model-based approach in classification. SpringerLink. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque non elit mauris. Cras euismod, metus ac finibus.
- Song, Y. (2015). Decision tree methods: applications for classification and prediction. PubMed Central (PMC). Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque non elit mauris. Cras euismod, metus ac finibus.
- Comparative analysis of PCA and LDA. (2011, June 1). IEEE Conference Publication | IEEE Xplore. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque non elit mauris. Cras euismod, metus ac finibus.

THANK YOU

Presented By : Sarah Ali Mohamed