

College Students Completing A Preliminary Year

Using logistic regression

Name: Likhil Naik Vislavath

The Issues

The project requires building a logistic model to predict the success or failure of college students completing a preliminary year. The data, available in both Excel and CSV formats, comprises 19 variables and a score of 1 for successful completion and 0 for failure. The aim is to create a predictive model that identifies the most useful variables for predicting the outcomes. The challenge is to identify which factors contribute significantly to the model's predictive power and which do not.

Findings

Discussion

Appendix A : Method

Appendix B: Results

Appendix C: Code

```
import pandas as pd  
import numpy as np
```

```
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.feature_selection import SelectFromModel
from sklearn.metrics import confusion_matrix, roc_auc_score, roc_curve
data = pd.read_excel("D:/MSDS/MTH522/assignm ent2/Preliminary college year.xlsx")

data.info()
data.isna().sum()
data.fillna(data.mean(), inplace=True)
X = data.drop(columns=['success_score'])
y = data['success_score']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=123)
logit = LogisticRegression(random_state=123)
logit.fit(X_train, y_train)
probabilities = logit.predict_proba(X_test)
```

References:

<https://chat.openai.com/chat>