

Introduction to Artificial Intelligence

Prof. Dr. Hamed Shariat Yazdi, Darmstadt University of Applied Sciences

Cross-review protocol

| | |
|-----------------|--|
| Assignment: | 03a. Machine Learning - Classification |
| Reviewed team: | Mo3y-Team-05 |
| Reviewing team: | Mo3y-Team-06 |
| Date: | |

Using the **median** instead of the **mean** to handle missing values is better when there are **outliers** in the data.

```
imputer = SimpleImputer(missing_values=np.nan, strategy='mean')  
X = pd.DataFrame(imputer.fit_transform(X), columns=features)  
test_data[features] = pd.DataFrame(imputer.transform(test_data[features]),  
columns=features)
```

We tried both and chose the mean in this case
