

1. Data

Number of Instances:

Class Label(s):

Features:

Training/Validation / % Testing %

2. Training

Model:

Hyperparameters:

Training Set Performance:

Validation

3. Evaluation

Class Balance:

Evaluation Metric(s):

Performance:

4. Your Conclusions

Was there any evidence of overfitting?

Was there any evidence of data leakage?

Was the evaluation metric appropriate?

Do you trust the validity of these results?