

Model Performance

Model Name: SupportVectorMachineOnLocation linear Test Date: 23/03/2022 15:56:42 Creator: Tobias Rothlin



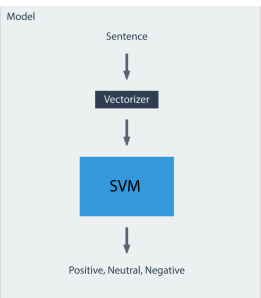
Overview

ML Principle:
Support Vector Machine

References:

- [Sentiment Analysis SVM](#)
- [Scikit SVM Kernels](#)
- [Scikit feature extraction](#)
- [Scikit Vectorizer](#)

Algorithm Description:
Support vector machines are a robust supervised learning model based on statistical learning. The idea is to find a Hyperplane separating the different classes with the most separation between the closest points. Before the SVM can classify a sentence, the sentence needs to be vectorised. To accomplish the Scikit learn, Tfidf Vectorizer is used. The Vectorizer converts the sentence to a fixed feature vector. With the vectorised sentences, the model can be trained. The best hyperplanes are found in the training step based on the training data. The flexibility of the hyperplane can be defined by the Kernel (linear, sigmoid, RBF). RBF is used for non-linear problems and is also a general-purpose kernel. This model uses a linear kernel.



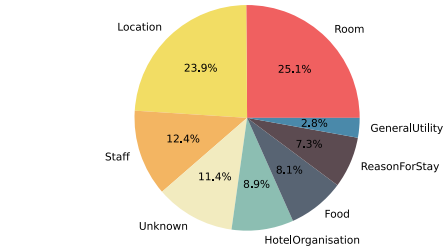
Classification Pipeline

Metrics

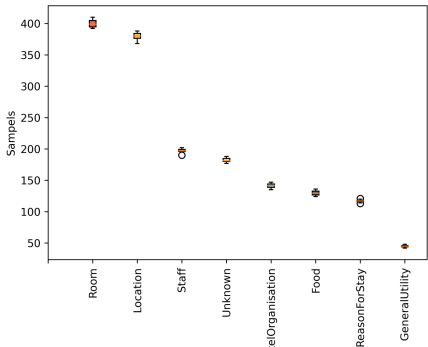
Data: ClassifiedDataSetV1.2 with 10 folds cross validation
Split seed: 4.83819

Training Dataset
(average)

| Classes | Number of samples |
|-------------------|-------------------|
| Room | 399 |
| Location | 379 |
| Staff | 197 |
| Unknown | 181 |
| HotelOrganisation | 141 |
| Food | 129 |
| ReasonForStay | 117 |
| GeneralUtility | 45 |



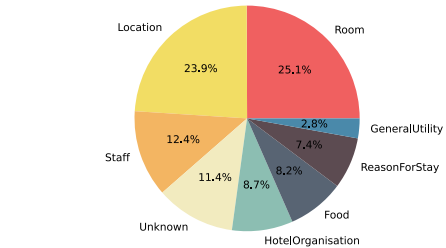
Average distribution of the samples



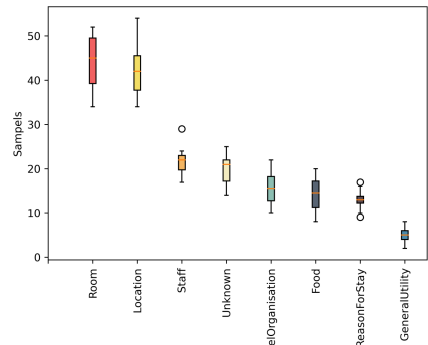
Distribution of the samples contained in each test split

Test Dataset
(average)

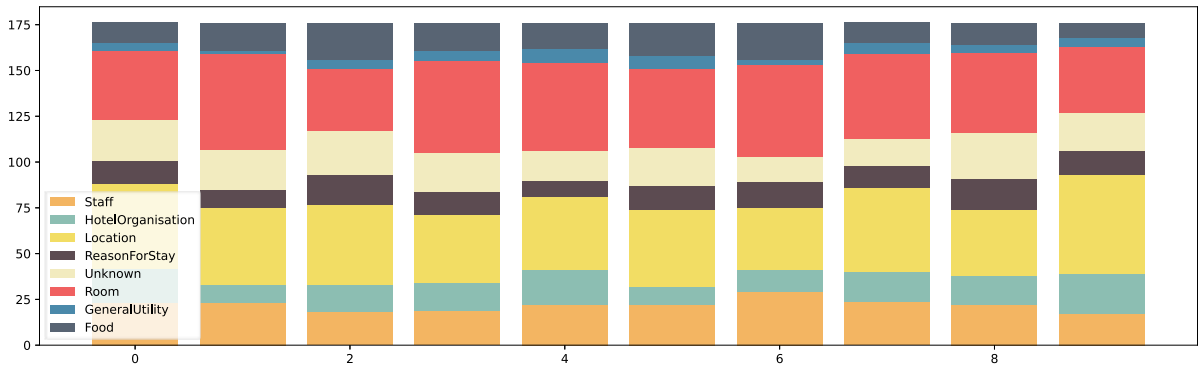
| Classes | Number of samples |
|-------------------|-------------------|
| Room | 44 |
| Location | 42 |
| Staff | 21 |
| Unknown | 20 |
| HotelOrganisation | 15 |
| Food | 14 |
| ReasonForStay | 13 |
| GeneralUtility | 5 |



Average distribution of the samples



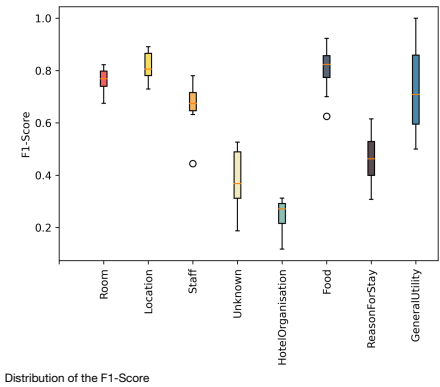
Distribution of the samples contained in each test split



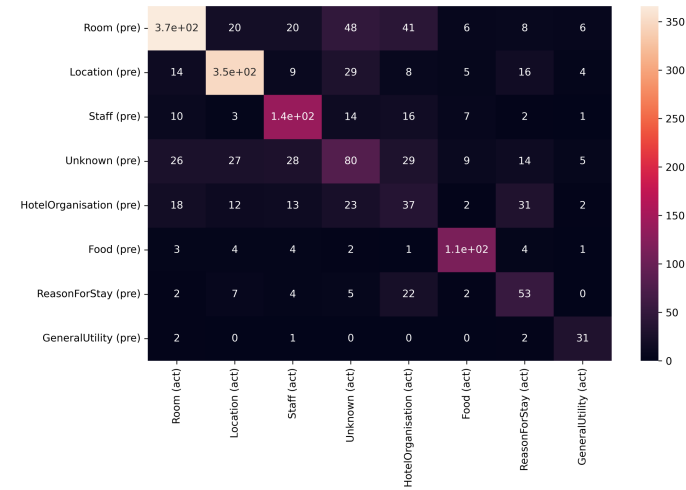
Detailed training split composition

Classification Performance

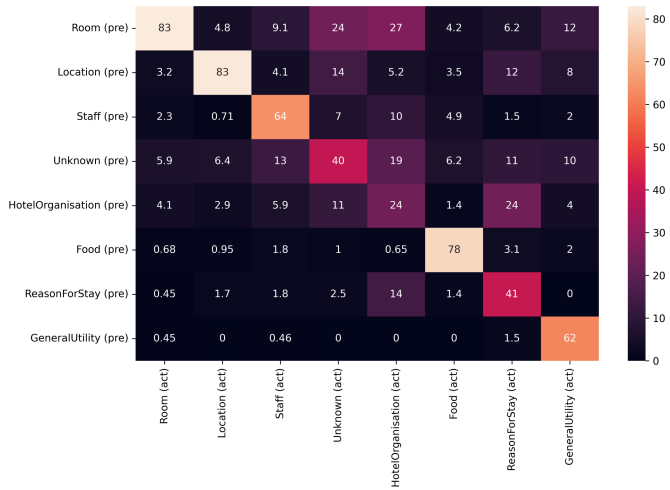
| Classes | Precision | Recall | F1 Score |
|-------------------|-----------|--------|----------|
| Room | 71.07% | 82.99% | 76.57% |
| Location | 80.37% | 82.66% | 81.50% |
| Staff | 72.54% | 63.93% | 67.96% |
| Unknown | 36.70% | 39.80% | 38.19% |
| HotelOrganisation | 26.81% | 24.03% | 25.34% |
| Food | 85.61% | 78.47% | 81.88% |
| ReasonForStay | 55.79% | 40.77% | 47.11% |
| GeneralUtility | 86.11% | 62.00% | 72.09% |
| Accuracy | | | 66.36% |
| Macro Average | 64.37% | 59.33% | 61.33% |
| Weighted Average | 66.17% | 66.36% | 65.94% |



ConfusionMatrix:



Normalised ConfusionMatrix:



F1 Socre by split:

