Model Performance

Model Name: SupportVectorMachineOnLocation rbf **Test Date:** 23/03/2022 15:59:06 **Creator:** Tobias Rothlin

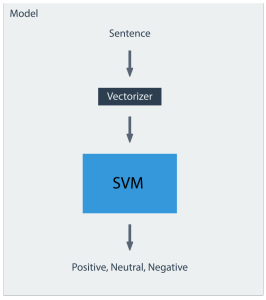
Overview

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 rbf kernel.



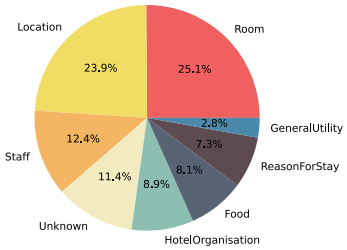
Classification Pipeline

Metrics

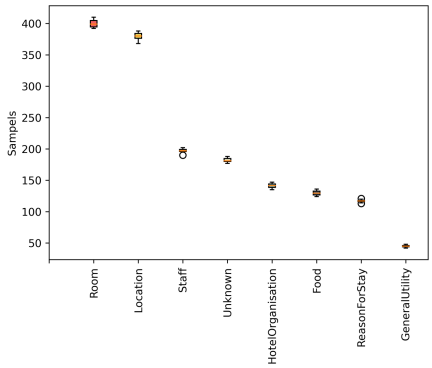
Data: ClassifiedDataSetV1.2 with 10 folds cross validation
Split seed: 4.83819

Training Dataset
(average)

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



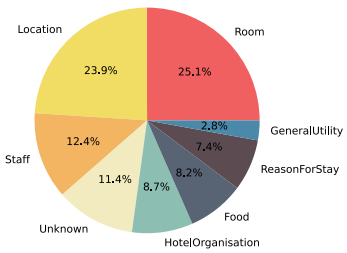
Average distribution of the samples



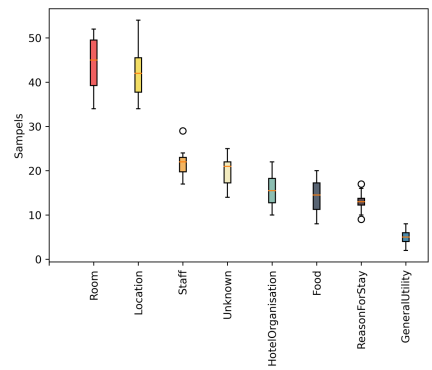
Distribution of the samples contained in each test split

Test Dataset
(average)

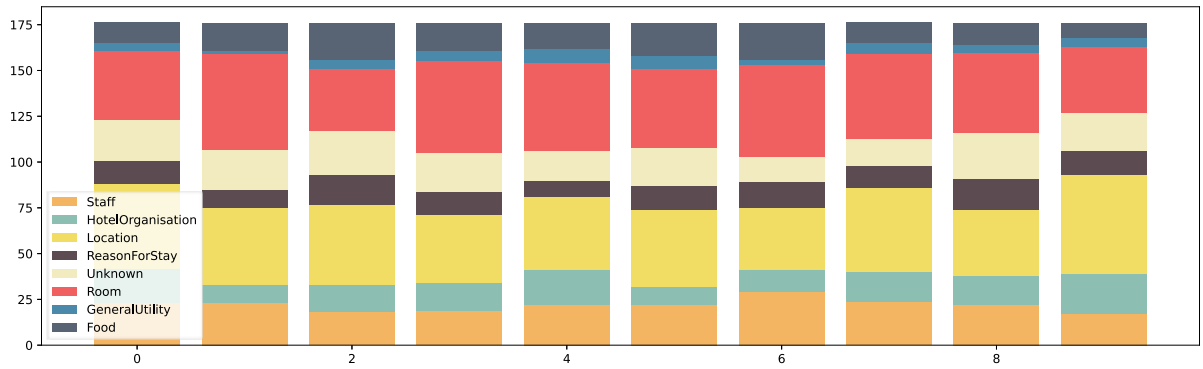
Classes	Number of samples
Room	44
Location	42
Staff	21
Unknown	20
HotelOrganisation	15
Food	14
ReasonForStay	13
GeneralUtility	5



Average distribution of the samples



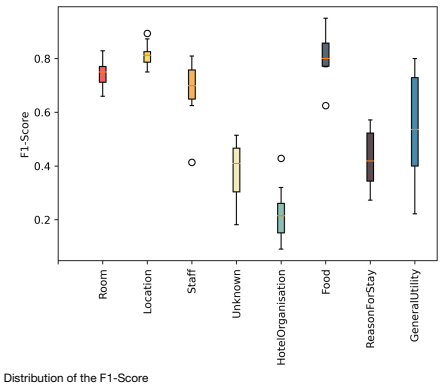
Distribution of the samples contained in each test split



Detailed training split composition

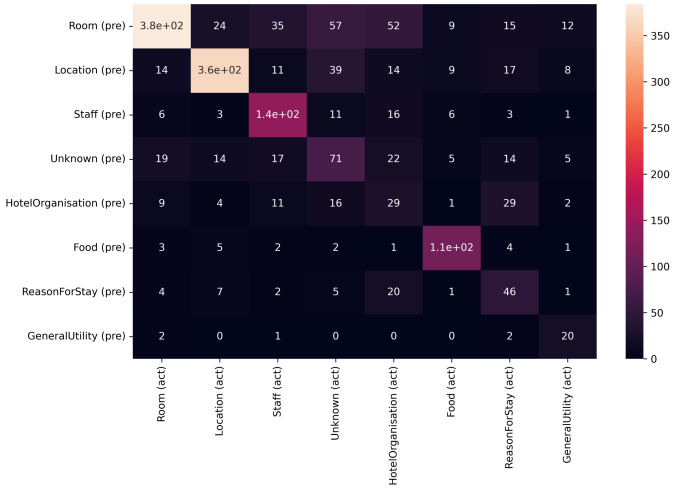
Classification Performance

Classes	Precision	Recall	F1 Score
Room	65.31%	87.07%	74.64%
Location	76.47%	86.46%	81.16%
Staff	75.27%	63.93%	69.14%
Unknown	42.51%	35.32%	38.59%
HotelOrganisation	28.71%	18.83%	22.75%
Food	86.26%	78.47%	82.18%
ReasonForStay	53.49%	35.38%	42.59%
GeneralUtility	80.00%	40.00%	53.33%
Accuracy			66.31%
Macro Average	63.50%	55.68%	58.05%
Weighted Average	64.67%	66.31%	64.50%

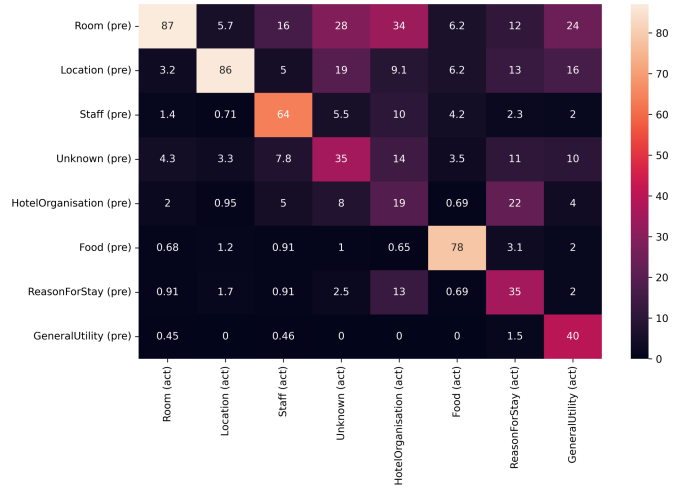


Distribution of the F1-Score

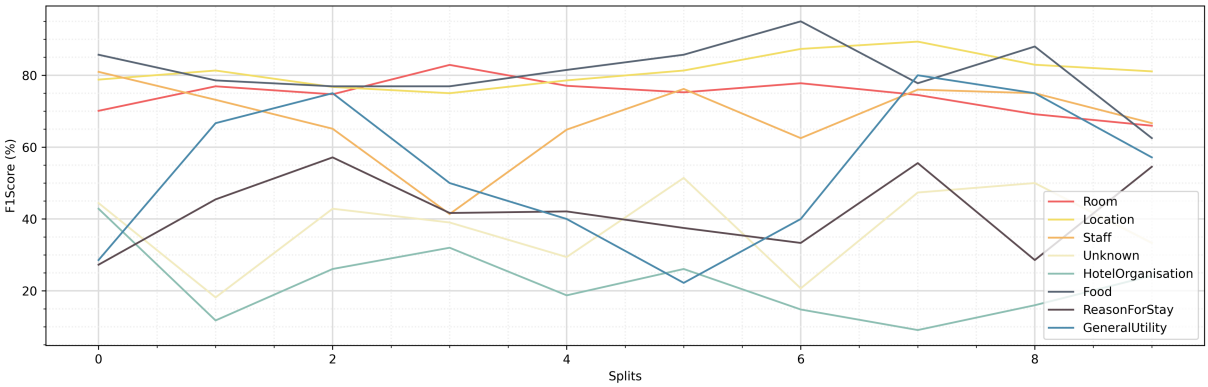
ConfusionMatrix:



Normalised ConfusionMatrix:



F1 Socre by split:



F1-Score per split