

Model Performance

Model Name: ScoreClassificationV1 Test Date: 11/03/2022 10:07:21 Creator: Tobias Rothlin

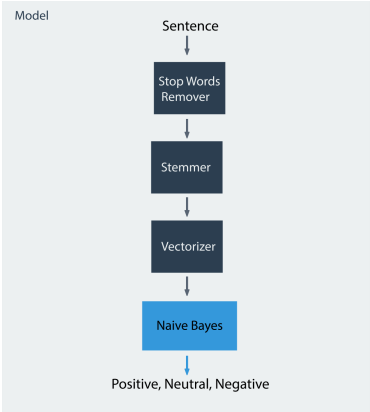


Overview

ML Principle:
Multinomial Naive Bayes

- References:**
- [MultinomialNB Explained](#)
 - [Stanford NLP Course](#)
 - [Stanford NLP Lecture](#)
 - [English Stopwords](#)

Algorithm Description:
The learning algorithm used in this classification is the Multinomial Naïve Bayes. This approach was chosen as it is easy to implement and is computational very efficient. The first step in the classification pipeline is removing all stop words for example 'i', 'me', 'my', 'myself', etc. A list of English stop word is provided by the nltk module. The stop words remover just removes every word that is in the list of stop words. Next the sentence is passed through the stemmer. Stemmers remove morphological affixes from words, leaving only the word stem. This is done with the PorterStemmer class from the nltk module. The final preprocessing step is to vectorize the sentence. This results in a bag of words representation of the sentence. First all the words must be tokenized and then counted. The result will be a numerical feature vector. To generate this vector the CountVectorizer class from sklearn is used. This class implements both tokenization and occurrence counting in a single class. With the sentence now represented in a vector the Naïve Bayes classifier can work with this vector. For the implementation of the Naïve Bayes classifier the MultinomialNB class (sklearn) is used.

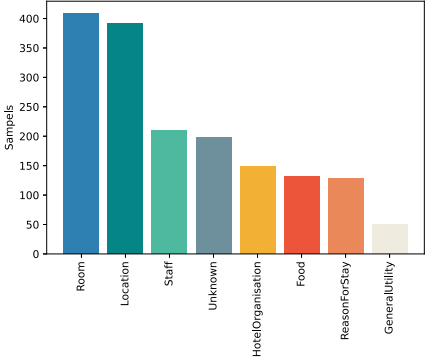
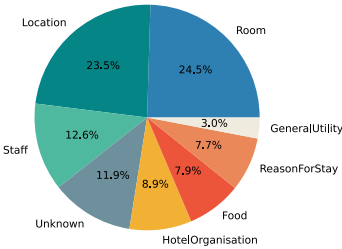


Classification Pipeline

Metrics

Training Dataset

Classes	Number of samples
Room	409
Location	392
Staff	210
Unknown	198
HotelOrganisation	149
Food	131
ReasonForStay	129
GeneralUtility	50



Classification Performance

Classes	Precision	Recall	F1 Score
Room	93.10344827586206%	77.14285714285715%	84.37500000000001%
Location	90.0%	84.375%	87.0967741935484%
Staff	100.0%	70.0%	82.35294117647058%
Unknown	25.0%	50.0%	33.33333333333333%
HotelOrganisation	12.5%	50.0%	20.0%
Food	78.57142857142857%	91.66666666666666%	84.61538461538461%
ReasonForStay	0.0%	0.0%	0%
GeneralUtility	0.0%	0.0%	0%

ConfusionMatrix:

