

Shahjalal University of Science and Technology
Department of Computer Science and Engineering,

CSE 476



Machine Learning Lab Final

Md. Jakir Hossain

Reg. No.: 2013331050

4th year, 1st Semester

Md. Abul Kashem

Reg. No.: 2013331068

4th year, 1st Semester

Department of Computer Science and Engineering

Submitted to,

Ayesha Tasnim

Assistant Professor

Department of Computer Science and Engineering

31st March, 2018

Fake News Detection

We collect the data of fake news detection from Arnob and Mridul. Their work is to determine a news fake or real from pre-determined real or fake news data.

Data Pre-process

Pre-process is performed by sklearn feature CountVectorizer, TfidfVectorizer and stop words. In both of Count Vector and TF-IDF Vector approach we split the dataset into training and testing dataset. Then predict the performance by applying different Supervised machine learning algorithm. We consider 75 % data for training and 25 % data for testing.

Input Feature: Value of Count vector and TF-IDF vector.

Output Class: Status of news (Fake or Real).

Artificial Neural Network

We have used 2 hidden layer which number of neuron are 35 and 41 and $\alpha=1e-5$.

Accuracy In TF-IDF approach= 87.1 %.

Accuracy In count vector approach= 91.7 %.

Naive Bayes

Accuracy In TF-IDF approach= 86.0 %.

Accuracy In count vector approach= 89.3 %.

Decision Tree

Accuracy In TF-IDF approach= 72.9 %.

Accuracy In count vector approach= 82.1 %.

K Nearest Neighbour

Accuracy In TF-IDF approach= 62.2 %.

Accuracy In count vector approach= 78.7 %.

Support Vector Machine: We used kernel='linear' and gamma=1 for SVM.

Accuracy In TF-IDF approach= 87.4 %.

Accuracy In count vector approach= 88.1 %.

Discussion

Our performance of fake news detection by artificial neural network is better than previous work. But other algorithm does not perform as well as ANN.

Apartment Price Prediction

We collect the data of apartment price prediction from Alif and Tuhin. The main work of their project is to predict the price of an apartment according to some given information of that apartment.

Data Pre-process

The data for apartment price prediction is almost pre-processed. We just normalize the data by removing two column named date and id. We consider 80 % data for training and 20 % data for testing.

Input Feature

Number of bedrooms, number of bathrooms, living sqft, lot sqft, floors, waterfront, view, condition, grade, sqft above, sqft basement, built year, zipcode etc.

Output Class

Price of apartment.

Linear Regression

We have used 2 hidden layer which number of neuron are 35 and 41 and $\alpha=1e-5$.

Accuracy = 71.56 %.

Gradient Boosting Regression

We have used learning rate=.1, number of estimators=100, maximum depth=5 for gradient boosting regression.

Accuracy= 88.84 %.

Despite these 2 approach we have also applied SVM, Decision tree, Naive Bayes, K nearest neighbour algorithm to predict apartment price. But their accuracy is too low for predicting apartment price.

Discussion

Our performance of apartment price prediction is not as good as previous work done by Alif and Tuhin. They use Random forest classifier for better performance.

Signature Detection

At first we collect the image data of Signature detection from Arnab Sen Sharma Api . Then we pre-process these data and apply 2 different machine learning algorithm.

Image Pre-process

At first we sharp the signature data, then we re-size each of the images into w*h (200 * 100) , after that we change the image each pixel value by dividing 255 . Then we change the images with gray scale. After that we convert each image into 200*100 size array *.txt format .We use this *.txt data set to apply machine learning algorithm. We distribute our data set 80% for training and 20% percent for testing purpose .

Input Feature

Pixel value in .txt format of each image.

Output Class

Numeric value of image.

Artificial Neural Network

For applying Artificial Neural Network (ANN) we have used **sklearn**. We have used 2 hidden layer which number of neuron are 35 and 41.

Accuracy= 89 %.

Support Vector Machine

For applying Support Vector Machine we have used **sklearn SVM** . Here we set kernel='rbf', c=4 and gamma=1.

Accuracy= 92 %.

Discussion

The accuracy of signature detection by Arnob and Mridul was 85 % which is less than our both applied approach.