QEEXO ML CHALLENGE

Proposed Solution:

Feature Selection:

We have audio files for the impact and some attributes relating to the touch, first these audio files were taken which are read using the scipy library in python giving output the samples and also there sample rate (256 samples). The main question comes how to get important features out of these audio files which actually will help to differentiate between the samples? The technique applied was **MFCC** which is a representation of the short-term power spectrum of a sound. It stands for, Mel Frequency Cepstral Coefficents (MFCCs), they are a feature widely used in automatic speech and speaker recognition. The steps involved in calculating MFCC are mentioned below:

1. Frame the signal into short frames.
2. For each frame calculate the periodogram estimate of the power spectrum.
3. Apply the mel filterbank to the power spectra, sum the energy in each filter.
4. Take the logarithm of all filterbank energies.
5. Take the DCT of the log filterbank energies.
6. Keep DCT coefficients 1-13, discard the rest.

A detailed explanation of MFCC is presented in [2].

MFCC was calculated using python library “Scikit-Talk box” whose files are included in the code files with a Readme.txt explaining the way to build it. In short use the command **python setup.py install**

As the signal had only 256 samples, MFCC for only one frame was calculated giving out 13 coefficients i.e. features for each sample. At the end we get a feature vector of 20,458x13.

Just to get an estimate how these features are working a SVM classifier (Implemented using Scikit learn library of python) was trained on these 13 feature input which approximately gave an accuracy of 92%. Giving a hunch that the features are fine and giving an okay result.

So next step would be to think how to increase the accuracy, this was done by now considering other data given i.e. the touch information. By manually checking out the CSV files there were 6 attributes defining touch, the coordinates(X,Y) , pressure orientation , major and minor axis. By just looking on the attributes we can estimate that coordinates would not help to classify touch, neither orientation and pressure as they were constant over all the samples(0 , -1). But major and minor axis of the ellipse forming a touch could be helpful as the touch between a pad and a knuckle would form different ellipses. So now these two features were also added to over old feature vector giving a total of 20,458x15.

This data was again tested, giving an accuracy of around 50% for default hyper parameters SVM. Conveying that those ellipses dimensions are indeed not helpful, we discard them getting back to the 13 feature array data. After that, features like filter banks energies which were obtained in between step of MFCC calculations were also taken as features, giving no improvement in results, other type of features like Delta (The MFCC feature vector describes only the power spectral envelope of a single frame, but it seems like audio would also have information in the dynamics i.e. what are the trajectories of the MFCC coefficients over time given by delta , a good explanation is given in [2]) giving no improvements in result.

Technique like basis expansion was also used to increase the features from 13 to 560, turns out it didn't increase accuracy so it was also discarded. A basis function expansion augments/replaces the attributes of a dataset with transformations of these attributes. For instance, given an input attribute X, a basis function expansion could map this attribute to three features: 1, X, X^2---a "polynomial basis." This mapping allows various learning algorithms and statistical procedures to capture nonlinear trends in the data while still using linear models to analyze these transformed attributes.

Future work:

In future more characteristics of sound can be extracted and tried upon and could even work in better way. Some of them may actually help are **wavelet transforms** like DWT, CWT, Wavelet packets, morlet wavelet or using **wigner-ville** distribution, **EMD**, **spectral kurtosis** and many more

Model Selection:

Now we have data with good features extracted, next step comes choosing a good model.

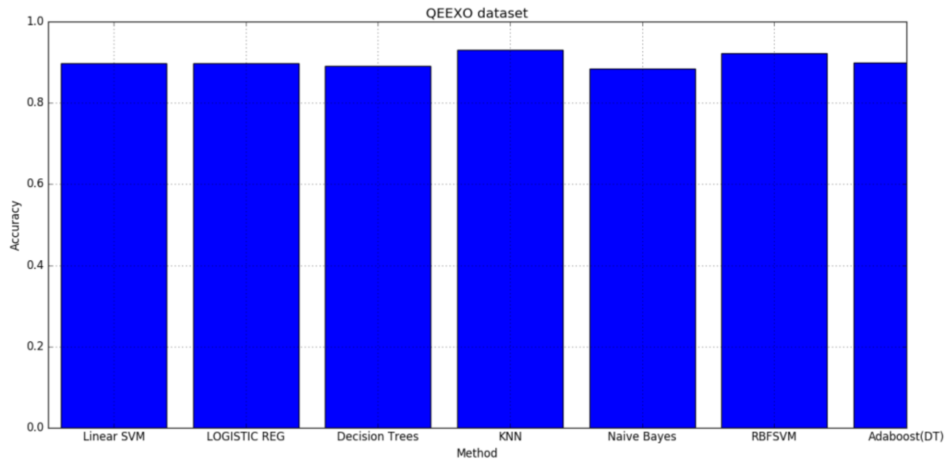
So about 6 different classifier models were chosen and trained upon they include, Naïve bayes, decision trees, SVM (Linear, RBF) , KNN, Logistic regression, Adaboost with decision trees with their default parameters. Their accuracies are compared and it turned out KNN worked out better of all.

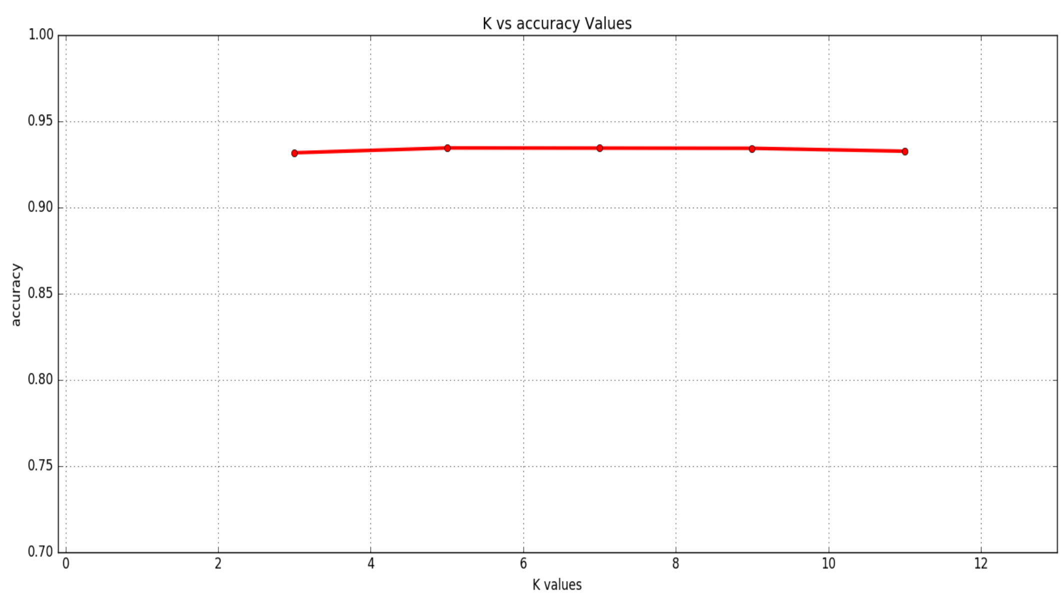
Cross Validation:

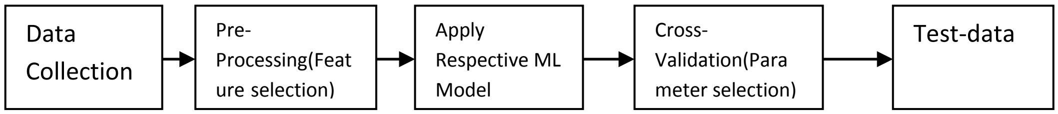
The best model was KNN , to avoid overfitting and even to find optimum hyper parameters 5 fold cross validation was done overall giving model accuracy of 93.5% with K=5 (Number of neighbors parameter). In *5*-fold cross-validation, the original sample is randomly partitioned into *5* equal sized subsamples. Of the *5* subsamples, a single subsample is retained as the validation data for testing the model, and the remaining *4* subsamples are used as training data. The cross-validation process is then repeated *5* times (the *folds*), with each of the *5* subsamples used exactly once as the validation data. The *5* results from the folds can then be averaged to produce a single estimation.

Test Data:

To end, this fitted model was applied to the test data giving predicted labels as output. The predicted test file is attached to the folder along with this report and python code (Jupyter notebook). Below, bar graphs shows the results of applying different models to the feature engineered data implemented using matplot library of python, line graph shows the accuracies with respect to different K values for KNN. Below block diagram shows the steps taken care of when designing this model.







Conclusion:

So a binary classifier was implemented for data containing about 13 important features using KNN with K=5, obtaining about 93.5% training accuracy, cross validation with 5 fold techniques was used to get optimal hyper parameters.

References:

[1] "TapSense: Enhancing Finger Interaction on Touch Surfaces" *Chris Harrison Julia Schwarz Scott E. Hudson*,Human-Computer Interaction Institute and Heinz College Center for the Future of Work;Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213

[2] http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/