Machine Learning for getting a phone number

Given data is phone number and scrambled letters. Feature that I have used is number of occurrences of the digit in both the cases. Linear regression is used as a classifier.

```
Training_Data -> [pre-processing] -> [Machine Learning Approach]

Test_Data -> [Machine Learning Approach] -> [post-processing] = telephone_number
```

Training data is given 100,000 samples of the numbers and scrambled letters. In post processing the training data set matrix is 100,000 by 15 matrix (As in one to 10 there are only 15 english syllables.)

The label dataset is 100,000 by 10 matrix having occurrences of digits 0 to 9 in the row.

This set the fed to the Linear regressor.

As a sample data set, A_Small dataset from google code jam is used.

Also, a python code without machine learning was written and checked for correctness on google code jam website. This output file is used for the verification.

As given data is too large, there is more time needed for the generation of the feature matrix. Hence, the predicted feature matrix is stored in a file and is accessed from separate python code for post processing and generating the phone number.

The output can be seen by running prediction.py file.

gettingPhoneNumber.py shows non machine learning and huristic based way (Tested for A large data set in google code jam).

gettingPhoneNumberML.py writes prediction to the file and uses machine learning to do it.