

# Classification Systems Developed to Predict Trail Outcome Using Transcribed Dialogue During a Trail

*Indumathi Desabathina*  
*School of Computing, University of Utah*  
*Salt Lake City, Utah*  
*Email: u1265976@utah.edu*  
*UID: u1265976*

## 1. Overview of the Project:

Old Bailey Proceedings Online project made available a fully searchable and digitized collection of approximately 197,000 trail proceedings from the year 1676 to 1772. The objective of this project is to develop classification system that can predict decision of trails present in the Old Bailey Proceedings Online project. The transcribed dialogue of a trail is available in three formats as highlighted in the Table 1. Along with the dialogue of the trail, defendant's age, defendant's gender, number of victims, victims' gender, offence category, and offence sub-category details are available to develop classification systems. These details from here on are referred as miscellaneous attributes.

**Table 1: Available formats of a dialogue of a trail**

Available formats of the digitized dialogue of the trails	Bag-of-words
	Glove
	Tf-idf

Using theoretical and practical understanding of various machine learning concepts and algorithms, six algorithms are selected and used to developed six different classification systems. Those six algorithms and accuracy achieved using each algorithm are listed in Table 2.

**Table 2: Algorithms used to develop classification systems and accuracy achieved**

Algorithm	Format of the dialogue used	Evaluation Accuracy
Average Perceptron	Bag-of-words	68.952 %
Margin Perceptron	Glove	62.857 %
Support Vector Machine (SVM)	Bag-of-words	72.876 %
Logistic Regression	Glove	66.400 %
Ensemble Method	Glove	62.628 %
Neural Network	Bag-of-words	67.085 %

Of all the six classification systems highlighted above, SVM achieved the highest accuracy. Theoretical understanding of SVM suggests that SVM has better generalization ability and lower risk of over-fitting the data and results achieved in this project also suggests that. However, it is also to be noted that there is lot of scope to implement improvements in all the six classification systems and accuracies might improve significantly.

## **2. Ideas Explored for this Project**

Following ideas were explored before starting the project:

- What are the best methods to pre-process the input data?
- What are the available techniques, methods or concepts that can be used to handle sparse data efficiently?
- Try to understand the importance of normalizing the data
- Study various resource to identify six best machine learning algorithms to build and complete this project
- Search and identify Python libraries that can assist and bring value to this project

## **3. Ideas borrowed CS 6350 Course**

- It has been taught in the course that normalization will play a crucial role in development of a classification system. When normalization was not implemented in this project, it has been observed that results were skewed. However, after implementing normalization results are not as skewed as observed previously.
- It has been taught that SVM performs structural risk minimization which will minimize the VC dimension. Minimizing the VC dimension should yield better generalization ability and lower risk of over-fitting. The results indicate that of all the algorithms implemented SVM has better accuracy.
- It is highly essential to have a complete and theoretical understanding of a machine learning algorithms before actual implementation of those algorithms to develop real-world classification systems. CS 6350 course has offered most comprehensive learnings of machine learning concepts.

## **4. Learning obtained from this project**

- It might be relatively straight forward to implement any algorithm to develop a classification system. However, without proper understanding of the theory of the machine learning algorithms, it is extremely difficult to get significantly good accuracies without overfitting the data
- Real-world projects like Old Bailey Proceedings do not offer clean and processed data. Development of classification systems is always associated with tedious and time-consuming data processing to obtain input features that are appropriate for the model development.
- There exists a huge scientific community and academia who are continuously working to develop libraries that assist in professional developing real-world classification systems. This project offered an opportunity to explore those libraries and construct classification system using those libraries such as keras, and TensorFlow.

## 5. Results and Summary

### Average Perceptron

For the classification system developed using average perceptron, bag-of-words format was used. The learning rate that used in average perceptron is 1. After constructing the average perceptron system using learning rate of 1, the system is trained and tested on bag-of-words. The accuracy achieved on training, evaluation, and testing are 77.502, 68.761, and 68.800, respectively.

Following development of the system mentioned above, there was an attempt made to develop another system using the miscellaneous attributes. But it was identified that majority of miscellaneous attributes require lot of processing. For example, attributes age contains strings or statements about age. Considering time constraint. Only miscellaneous attributes that do not require processing are embedded into the new system such as number of victims. Using the same learning rate, the system is trained, evaluated and tested. ***Accuracies achieved in each stage are 77.66, 68.444, and 68.952.***

### Margin perceptron

Using glove format, a margin perceptron classification system is developed. Along with the glove, 5 out of 7 miscellaneous attributes are used in development. To add these 5 miscellaneous attributes to the system lot of processing is done. Examples of processing the attributes include:

1. Converting object into integers for defendant's age
2. Label encodings for the remaining categorical features
3. And finally normalizing all the 5 miscellaneous attributes.

With a total of 305 feature set, which include 300 from the glove and 5 from the miscellaneous attributes, margin perceptron algorithm with  $\mu = 0.01$ , and learning rate = 0.1 is applied to derive a classification system. ***The accuracies achieved in training, evaluation, and testing are 64.6411, 62.857, and 62.133.***

### Neural Network

Using the same input features used in developing margin perceptron system a neural network is developed. In the neural network, three hidden layers are used, and each layer is of size 100, 50, and, and 25, respectively. The activation function used in the hidden layer is Rectified linear unit (ReLU), and the activation function used in the output layer is sigmoid. ***The accuracies achieved in training, evaluation, and testing are 88.4971, 67.085, and 67.111.***

### Support Vector Machine (SVM)

Implemented SVM on bag-of-words format and developed the classification system. SVM is selected because theoretical framework of SVM suggests that it has better generalization ability and lower risk of over-fitting the data. To further improve the performance of the SVM 5-fold cross validation is performed. Cross validation is intended to identify optimal hyper parameters. Optimal hyper parameters will most likely give high accuracy without overfitting the data. After

5-fold cross validation, the *accuracies achieved in training, testing, and validation are 80.6971, 72.0444, and 72.876, respectively.*

### **Logistic Regression**

Glove format was selected to build a classification system using logistic regression. Five-fold cross validation is performed on this system to determine hyperparameters, which are learning rate and trade off as 0.001 and 10,000. The accuracies achieved in training, testing, evaluation phases are 65.628%, 65.022%, and 66.400% respectively.

### **Ensemble Method**

In the classification system developed using ensemble method, perceptron is used as weak learner and SVM as the strong learner. To determine hyperparameters for the weak learner, five-fold cross validation is performed. Using those hyperparameters, 50 perceptron (weak learners) are built. Using 50 perceptron, input feature set is transformed. On the transformed input feature set strong learner (SVM) is applied. The accuracies achieved using Ensemble method in training, testing, and evaluation phases are 60.062%, 62.266%, and 62.628% respectively.

## **6. Future Work**

Future work that can be built on this project are as follows:

1. Extensive data exploration can be performed on that data set such as correlation analysis. Correlation analysis will assist in identifying the variables that exhibit poor strength out output variables. Dropping those variables from the model will improve the computational efficiency and accuracy.
2. Analyze and identify if multicollinearity exists in the independent variables. And if multicollinearity exists determine and implement right methods to handle multicollinearity.
3. Explore and determine right dimensionality reduction techniques that are appropriate to the dataset used in this project.
4. Developing a classification system using decision tree took relatively longer time as each attribute of miscellaneous variables has multiple values. Implementing ID3 with proper binning might yield better accuracy.
5. This report lacks vivid graphs and pictures that can clearly convey the results to the audience. A future work should develop visualization that can elaborately explains results achieved with each model and how and why each technique used in this project affected the result.