**Assignment 5**

**Computational Linguistics**

**Team members:**

**Burde Prerana Kamath**

**Ajit Balaga**

**Arpitha Kashyap**

**Question 1:**

a) Eager - lazy learning

Eager Learning:

* Eager learning methods construct generalized, input independent models with the training set and uses that on new data (test set)
* Even a slight change in the problem domain causes the model to classify incorrectly
* Doesn’t involve time consuming search for similar data
* The model can be build using the subset of the data as well
* The model is built to solve a single predefined problem
* If test data has unseen classes, the model will not be able to classify.

Lazy Learning:

* The Lazy learning method simply stores the data. Generalizing beyond this data is postponed until it receives new data (a test set)
* When it receives new data, the new examples are stored in the memory along with the earlier examples so that it can be used for classification. Thus, this technique can adapt itself to changes in problem domain
* It takes less time in training but the search for similar data is time consuming
* This method can be used for multiple problem-solving purposes
* High memory is required for storage as the learner often must deal with increasing amounts of data
* This method can deal with incomplete problem domain. Since it learns from new examples, it can classify unknown classes as well.

b) Generalization – abstraction

Generalization:

* A generalization operation defines similarities between objects. This operation does not affect the object’s representation. This implies no loss of information
* Abstraction reduces the number of properties
* Induction generally introduces it.

Abstraction:

* An abstraction operation changes the representation of an object by hiding or removing less critical details while preserving desirable properties. This implies loss of information
* Generalization adds new properties
* It allows deductions to be made.

c) Support Vector Machines - memory-based learning

Support Vector Machines

* Support vector machines are supervised learning algorithms that analyse data used for classification and regression analysis.
* SVM work well with outliers as it uses only relevant points to construct the support vectors
* SVMs have good performance in high dimensional data.

Memory-based learning

* Like lazy learning, Memory based learning will compare new test instances with the existing instances that are stored in memory
* Outliers affect memory based learning
* The performance decreases with increase in dimensions. Dimensional reduction must be done to avoid the cures of dimensionality

**Question 2:**

The tnt software was used to generate the POS tags for the feature model. The nouns, verbs and adjectives are used as feature models used to generate binary tweet labels, 0 and 1.

The timbl software was run to build the model and the screenshot of the output is given below.