**Speech Deterioration in Alzheimer’s disease: Case Study of Agatha Christie**

Statistics show that signs of Alzheimer’s in Agatha Christie can be seen in her works from her 70’s. So I take the works from the year 1950 as training set (36 works) and before 1950 (41 works )as testing set. The training dataset is used to identify the best machine learning model for our problem and the testing dataset is used to verify that model.

**Steps followed :-**

1. Collection of texts of Agatha Christie
2. Removing the punctuations and converting it to lower case(cleaning data)
3. Un-stemmed document was used
4. Features extraction
5. Analysis using Weka
6. Used java for coding

**Lexical features used:-**

1. Unique word count ( using Hashmap for this)
2. Frequency of nouns (POS tag NN and NNS)
3. Frequency of pronouns(POS tag PRP)
4. Frequency of conjunction(denoting coordinated sentences POS tag CC)
5. Frequency of Indefinite terms(thing, something, stuff)
6. Reduced sentences (POS tags VBG and VBN)
7. Interjections (POS tag UH)
8. Total word count
9. Count of Adjectives(POS tag JJ)
10. Count of verbs(POs Tags VB,VBD,VBG,VBN,VBP,VBZ)
11. Total sentences count
12. Normalized sentences count
13. Noun to verb ratio
14. Adverb
15. TTR
16. Brunet’s index
17. Honore’s Statistic (Higher value signifies richer vocab)
18. Mean length of words
19. Lexical density (ratio of words providing meaning to the text like nouns, adjectives, verb and adverbs to the total number of words, Lexical words give a text its meaning and provide information regarding what the text is about. More precisely, lexical words are simply nouns, adjectives, verbs, and adverbs. Nouns tell us the subject, adjectives tell us more about the subject, verbs tell us what they do, and adverbs tell us how they do it.)
20. Function words(Non- informative words in the text)
21. Use of punctuations
22. Year

**Machine Learning Models Used in the Regression type experiment**

1. Linear Regression
2. Regression Tree
3. K-nearest neighbors
4. Simple Linear Regression
5. Gaussian Processes
6. Kstar
7. Multilayer perceptron
8. Locally weighted Learning (LWL)
9. Additive regression
10. Bagging
11. CV Parameter Selection
12. Random Committee
13. Random sub space
14. Regression by Discretization
15. Stacking
16. Vote
17. Weighted Instances Handler Wrapper
18. Decision table
19. M5 Rules
20. ZeroR
21. Decision stump
22. Random Forest
23. Random Tree