**Predicting Grade Level of Educational Resources**

Presented By

**Dimpi Saikia 15CS60R08 Gowtham Nayak 15CS60R22 Kalyani Roy 15CS60R20 Survi Makharia 15CS60R01**

Dr. Under Plaban the of

**Bhowmick Guidance**

**Contents**

**▪ Introduction**

**▪ Framework**

**▪ Feature Selection**

**▪ Training**

**▪ Classification**

**▪ Milestones**

**Introduction**

▪ We want to target people of all grade labels, so we need to show them the relevant documents that they can understand.

▪ To do this, we require an automated document classifier that can classify documents easily.

▪ We are demonstrating an approach to predict the grade labels of documents.

▪ Here we are building a predictive grade label classifier that predicts the probability of belonging of a document to the grades specified in the training.

**Framework**

▪ A supervised training corpora containing the correct label for each input.

▪ In our model we are considering NCERT textbooks of different grades.

**Feature Selection**

We need to identify features of data that are salient for classifying, till now we have used three different features:

▪ Unigram probability

▪ Average word length

▪ Average sentence length

**Training Data**

▪ We decided to focus on textbooks that are suggested on reading lists at different grade levels in NCERT.

▪ This gave us the large amount of text we needed for building language models, and additionally, labeled data was readily available.

▪ We have considered data of 6-10 grade labels.

**Classification**

▪ We are doing probabilistic classification of a document into various grade labels

▪ The classifier that we are using now is logistic regression.

▪ But we will also use Naive Bayes, Random Forest, Support Vector Machine and then compare their performance.

**Milestones**

▪ Different encoding of pdf documents in the training corpus.

▪ For documents that contains words most of which are unseen, prediction is not significantly good.

▪ Some of the words in pdf are not separated by space, which leads to faulty unigrams.

**References**

• Martin, James H., and Daniel Jurafsky. "Speech and language processing."International Edition (2000).

• http://www.nltk.org/book\_1ed

• http://nlp.stanford.edu/courses/cs224n/2008/reports/12.pdf

• Language processing in e-learning, lecture slides by dr. plaban Bhowmick

**Thank You!**