PROJECT PROPOSAL

BBC NEWS CLASSIFICATION

OR – 610
DEEP LEARNING

Instructor: DR. VADIM SOKOLOV

BY
MAYANK DUBEY
(G01220775)

INTRODUCTION:

Compiler is a computer program which translates source code which is written in programming language to another computer language, however those are just a set rules not an intelligent machine which can take decisions. Things become interesting when human communicates and there are no rules or less rules, analyzing human language and giving linguistic power to the machine is called "Natural language processing".

"Natural Language Processing is the branch of AI which lets computer to understand, interpret and manipulate the text data". I am fascinated about how a machine can work on unstructured data and can make decision. The NLP is not new, it is started in 1950 when Alan Turing published the article called "Machine and Intelligence", since then it is growing and is important part of A.I.

GOAL

The main idea is given the news report the model will be able to classify report. It is the multiclass classification problem It has five classes they are business, entertainment, politics, sports and tech. The dataset is evenly distributed among 5 classes roughly has 500 data points for every class.

METHOD:

- **Machine Learning:**
  - ➢ Naïve Bayes
  - ➢ Random Forest
  - ➢ SVM
- **Deep Learning:**
  - ➢ Neural Networks
  - ➢ RNN
  - ➢ LSTM
  - ➢ GRU

TEXT SUMMARIZATION:

If there is enough time, I would also like to do text summarization. With every report the text summary is given.

DATASET:

Sharif, P. (2018, May 06). BBC News Summary. Retrieved October 09, 2020, from https://www.kaggle.com/pariza/bbc-news-summary/