PROJECT REPORT

# Hierarchical Attention Networks for Text and Document Classification

Guide: K. Ashesh

|  |  |
| --- | --- |
| 180030120 | Vasavi Gannina |
| 180030887 | Neerukonda Sai Sruthi |
| 180030956 | Chekuri Chekitha |
| 180031226 | Hari Priya Puppala |

**Abstract**

Evaluating all text data manually is time-consuming and laborious. A more efficient way to extract important information is to classify the text.

Text classification is a fundamental task in natural language processing. The goal is to assign unstructured documents (e.g., reviews, emails, posts, website contents, etc.) to one or more categories. These categories can be review ratings, such as star ratings, spam rating versus non-spam rating, or topic rating.

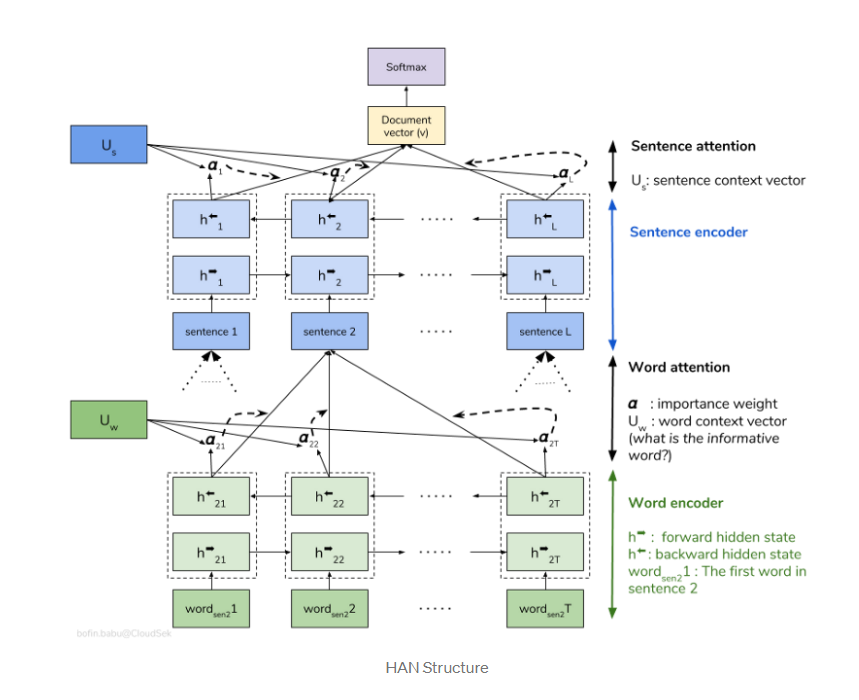
Basically, text classification can be used when there are certain tags to map to a large amount of text data. To know how to classify, we need to construct the classifiers which are obtained from the sorted data. In this way, the process of checking information becomes automated and therefore simpler. We use Hierarchical attention networks for text and document classification.

Hierarchical attention network uses word-level stacked recurrent neural networks followed by an attention model to extract such words important to the meaning of the sentence and assemble the representation of those useful words to form a sentence vector, then the same procedure is applied to the derived sentence vectors which then generates a vector that visualizes the meaning of the given document, and this vector can be bypassed to classify Text.

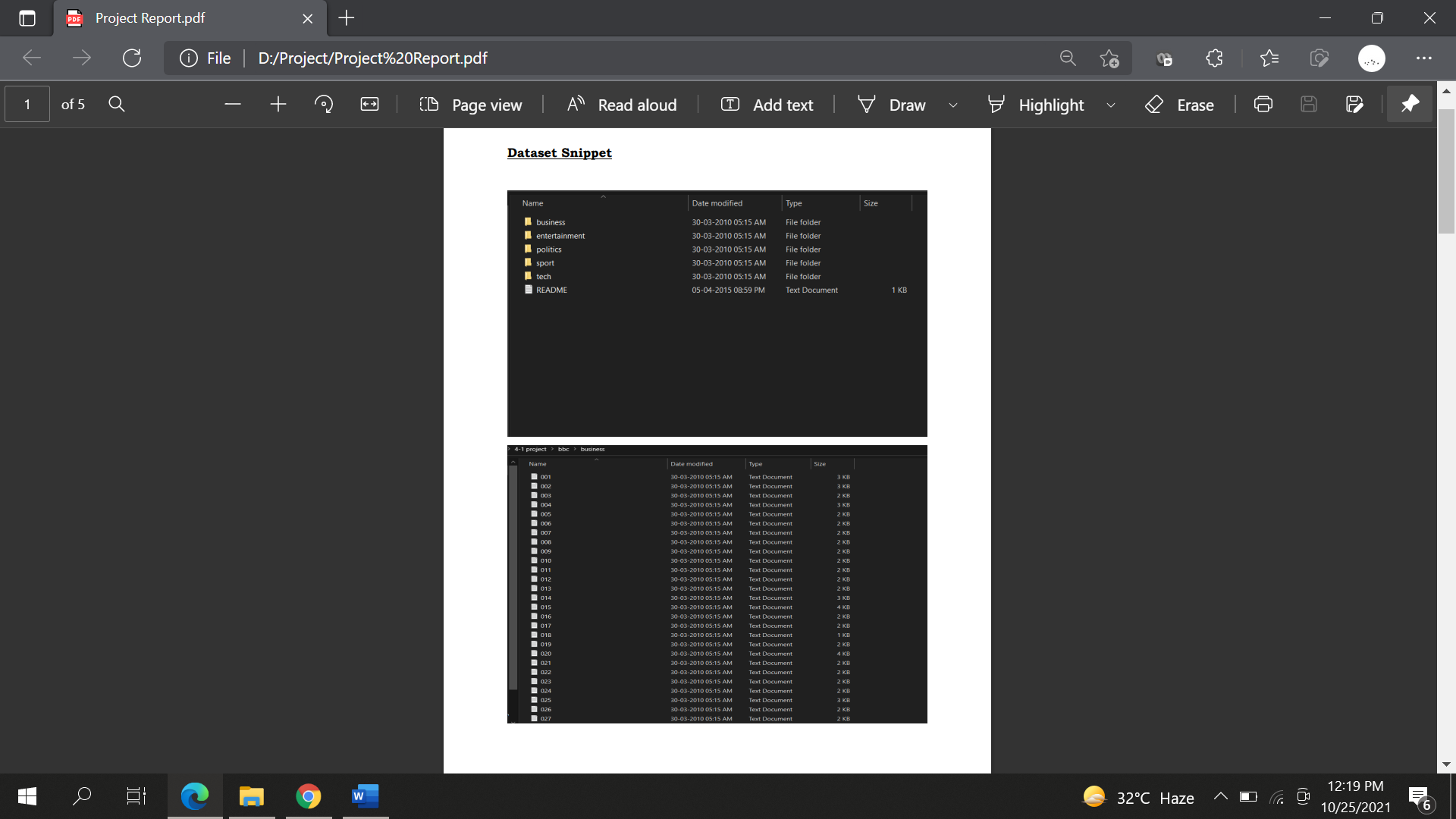
Bidirectional RNN and attention networks are the two parts that constitute the attention model. The bidirectional RNN returns a vector for each word by learning the meaning behind this word sequence. The attention network uses its shallow network to obtain corresponding weights for each word vector. Furthermore, the weighted sum of each vector is calculated. This weighted sum captures the entire sentence. The same procedure is applied to sentence vectors so that the final vector can capture the essence of the entire document. Since it works in two levels as an attention model, it is called hierarchical attention networks.

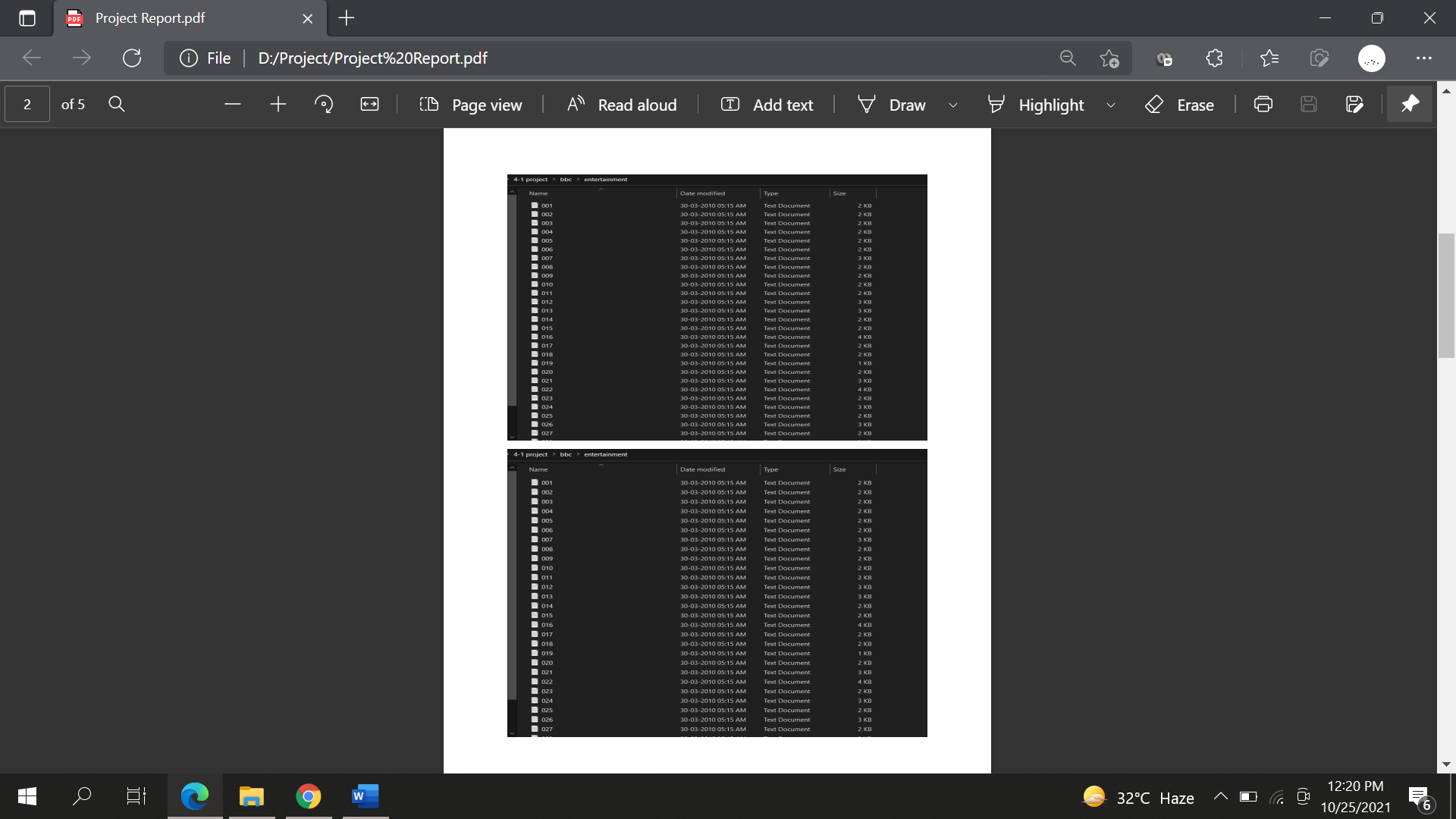
**Introduction**

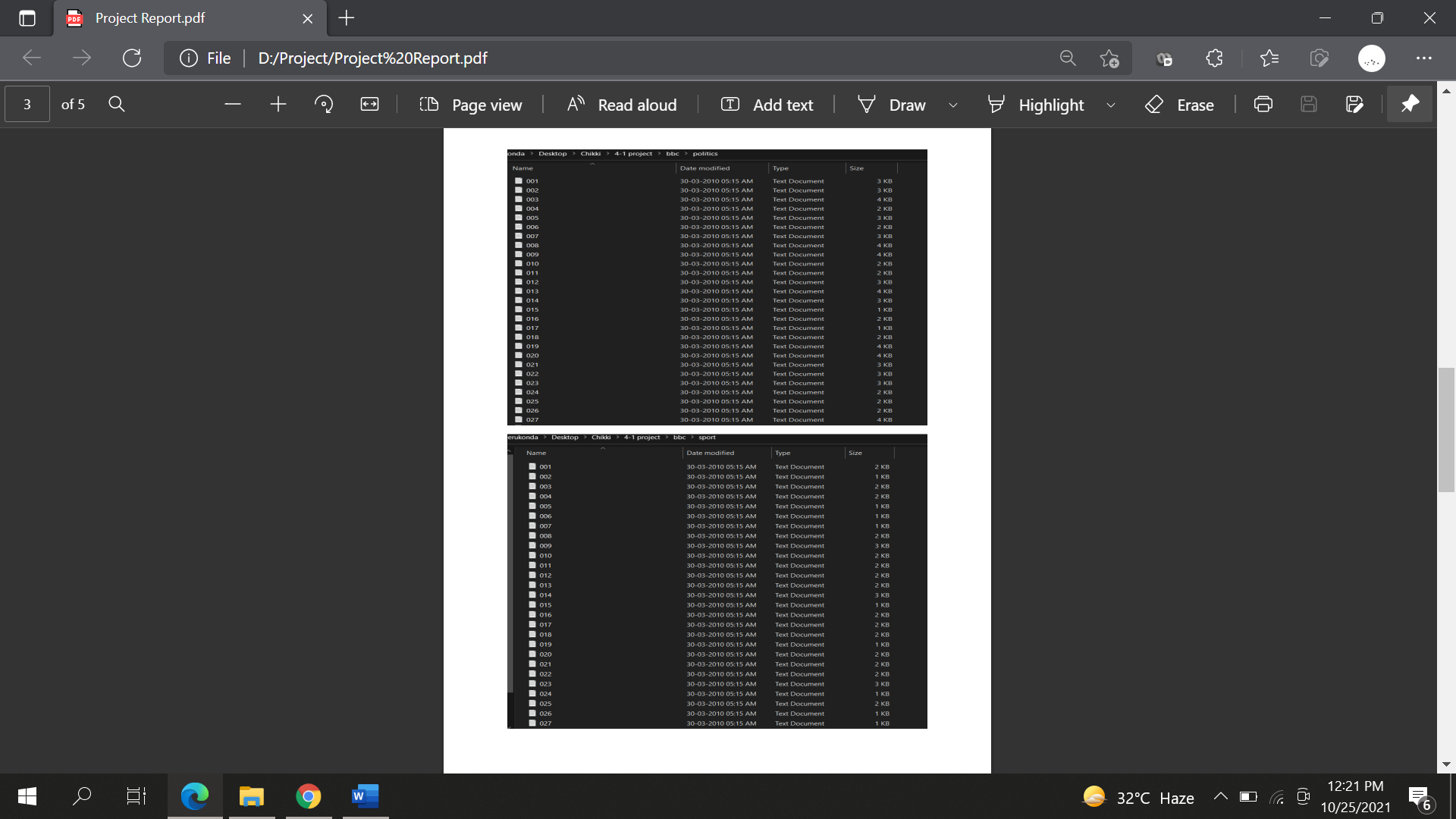
It uses stacked recurrent neural networks on word level followed by attention model to extract such words that are important to the meaning of the sentence and aggregate the representation of those informative words to form a sentence vector. Then the same procedure applied to the derived sentence vectors which then generate a vector who conceives the meaning of the given document and that vector can be passed further for text classification. The idea behind the paper is “Words make sentences and sentences make documents”. The intent is to derive sentence meaning from the words and then derive the meaning of the document from those sentences. But not all words are equally important. Some of them characterize a sentence more than others. Therefore, we use the attention model so that sentence vector can have more attention on “important” words. Attention model consists of two parts: Bidirectional RNN and Attention networks. While bidirectional RNN learns the meaning behind those sequence of words and returns vector corresponding to each word, Attention network gets weights corresponding to each word vector using its own shallow neural network. Then it aggregates the representation of those words to form a sentence vector i.e it calculates the weighted sum of every vector. This weighted sum embodies the whole sentence. The same procedure applies to sentence vectors so that the final vector embodies the gist of the whole document. Since it has two levels of attention model, therefore, it is called hierarchical attention networks.

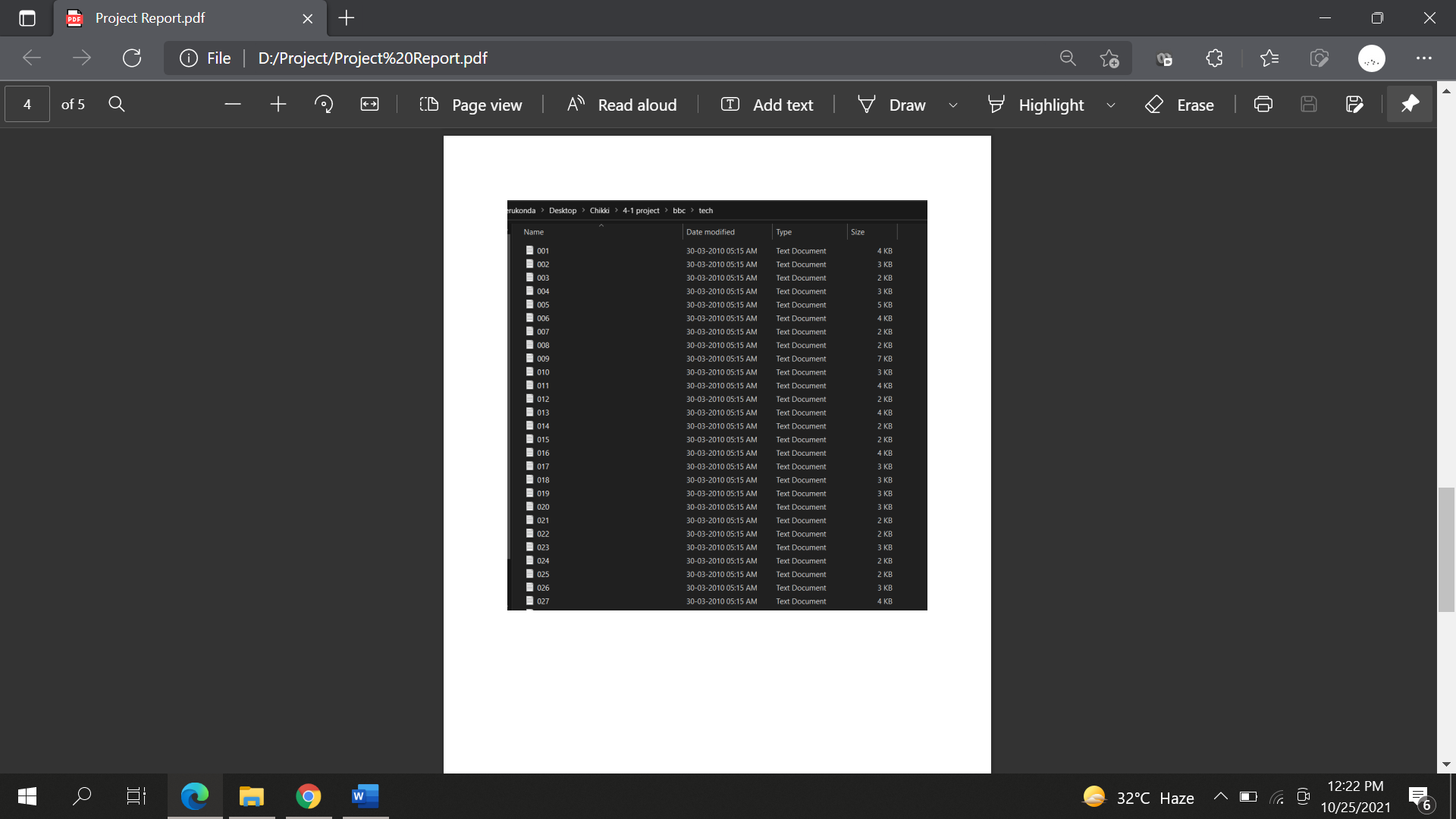


**Dataset Snippet**









**Code Snippets**

