Advanced Text Document Classification Techniques: Enhancing Accuracy and Efficiency in Real-World Applications

Bachelorarbeit

zur Erlangung des akademischen Grades

Bachelor of Science in Engineering (B.Sc.)

Eingereicht bei:

Fachhochschule Kufstein Tirol Bildungs GmbH

Web Business & Technology

Verfasser/in:

Jesse Lang

2110653142

Gutachter : Prof. (FH) Dr. Michael Kohlegger

Abgabedatum:

19. May 2024

Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Bachelorarbeit selbständig und ohne fremde Hilfe verfasst und in der Bearbeitung und Abfassung keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate als solche gekennzeichnet habe. Die vorliegende Bachelorarbeit wurde noch nicht anderweitig für Prüfungszwecke vorgelegt.

Kufstein, 19. May 2024

Jesse Lang

Sperrvermerk

Ich habe die Sperrung meiner Bachelorarbeit beantragt, welche von der Studiengangsleitung genehmigt wurde.

Kufstein, 19. May 2024

Jesse Lang

Contents

List of Figures

List of Tables

List of Listings

List of Acronyms

HTML HyperText Markup Language

JS JavaScript

FH Kufstein Tirol

Web Business & Technology

Abstract of the thesis: **Advanced Text Document Classification Techniques: Enhancing Accuracy and Efficiency in Real-World Applications**

Author: Jesse Lang

First reviewer: Prof. (FH) Dr. Michael Kohlegger

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan

semper.

19. May 2024

FH Kufstein Tirol
Web Business & Technology

Kurzfassung der Bachelorarbeit: **Advanced Text Document Classification Techniques: Enhancing Accuracy and Efficiency in Real-World Applications**

Verfasser: Jesse Lang
Gutachter: Prof. (FH) Dr. Michael Kohlegger

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan

semper.

19. May 2024

1. Introduction

The classification of text documents is essential across various applications. It demands a high accuracy and efficiency, in order to be beneficial. Traditional approaches often struggle with the diversity and complexity of modern textual data, emphasising the need for advanced techniques. Leveraging natural language processing (NLP), data mining and machine learning (ML) techniques holds promise in overcoming these challenges by enabling automated categorisation of textual documents based on their content, context and semantics. These procedures contain numerous obstacles, including defining accurate annotation of documents, the implementation of dimensionality reduction techniques to address complexities, utilizing appropriate classifier functions to obtain robust generalisation while avoiding overfitting (Aurangzeb et al., 2010).

One of the main sources for textual documents these days is the web, the amount that is available to us is constantly increasing. Unstructured textual formats, such as reports, emails, opinions, and news stories, are thought to contain approximately 80% or more of an organization's information. Studies indicate that unstructured formats contain almost 90% of the world's data. There's a clear necessity for the automatic extraction of valuable insights from vast amounts of textual data to aid human analysis (Aurangzeb et al., 2010).

This study aims to investigate how these advanced classification techniques, can enhance the accuracy and efficiency of categorizing diverse textual documents. Libraries, such as scikit-learn (Scikit-Learn, 2024), NLTK (NLTK, 2024) and Keras (Keras, 2024) and their broad spectrum of algorithms will be used to implement different approaches and exploring the interactions between these techniques. With the use of a BBC text document dataset with a ".csv" format does the research seeks to gain insights into optimizing classification performance in real-world applications.

The following research question arises from the objective: "How can advanced text document classification techniques, incorporating machine learning and natural language processing, be effectively employed to improve the accuracy and efficiency of categorizing diverse textual documents, and what factors influence the performance of such classification models in real- world applications?"

1.1 Motivation

The current state of research in text document classification techniques, with ML and NLP, showcases significant progresses towards enhancing efficiency and accuracy. Recent developments have seen the emergence of sophisticated deep learning architectures, such as CNNs and RNNs, which excel in processing complex textual data. Furthermore, has the integration of pre-trained language models like GPT and BERT revolutionized feature representation, allowing models to capture difficult semantic nuances. This dynamic landscape reflects ongoing efforts to refine classification models, making them increasingly adept at real-world applications.

2. Theory

The theoretical basis of this thesis provides the necessary background to comprehend the research discussed in the following chapters. Through an insight of relevant development in the field, various models, preprocessing methodologies, and evaluation metrics, this chapter aims to construct a theoretical framework for the thesis. Its objective is to empower the reader with the requisite understanding to contextualize the research outcomes and their significance.

2.1 Natural Language Processing

Natural Language Processing (NLP) is a branch of artificial intelligence and computer science, that deals with the interaction between computers and human language. It aims to enable machines to generate human language, process, and understand it. By employing a variety of techniques and different approaches, such as deep learning, rule-based systems, and statistical methods, is NLP capable of tackling different language-related tasks. The usage of it can be found in numerous areas, including machine translation, chatbots, text classification, and speech recognition (Helland, 2023).

2.2 Machine Learning

Machine learning refers to the development of computer programs that are learning from experience to complete and solve a certain task. The measurement of the performance is calculated by the ability to do so. A training dataset represents the "experience" that is acquired by machine learning models, which contains output and input pairs. From the analysis of these examples can the model recognize and generalise the patterns to new, unseen data. To simplify, can it be reflected to the process of human learning and adaption to new scenarios. Common applications include fraud detection, self-driving cars and personalised recommendations (Helland, 2023).

2.3 Deep Learning

Deep learning is a method that uses non-linear modules to transform the data at multiple levels of abstraction, allowing models to find patterns in the raw data. This suggests that deep learning models are discovering and learning different data features without the need for human interaction. Because of its universal learning, generalization potential, robustness, and scalability advantages, this can be used in a variety of applications without the need for precise feature engineering (Helland, 2023).

2.4 Evaluation Metrics

A machine learning model's performance is assessed using metrics. They are employed to evaluate the accuracy of the made predictions, to analyze the output of various models, and to fine-tune them for optimal performance. Different types of machine learning issues have various types of metrics available. The model selection process, the optimization procedure, and the overall understanding of the model's capabilities can all be impacted by the metrics chosen. Selecting the incorrect metrics might also result in a biased model, which aligns differently with the project's objectives (Helland, 2023).

2.4.1 Accuracy

One popular assessment metric for classification problems is accuracy. Out of all the samples in the prediction, it calculates the proportion of correctly classified samples.

Accuracy = Number of correctly classified instances/ Total number of instances

For balanced datasets, accuracy works well since it presents a realistic picture of the model's capabilities and performance. When datasets are unbalanced, accuracy might be misleading and more difficult to interpret. This may be due to a single label in the dataset that accounts for the majority of the samples; therefore, reasonable accuracy can still be obtained by projecting all samples to the dominant label. This does not imply that the model is good because it ignores the less common but no less significant labels. Accuracy in multi-label classification only takes into account samples where every label is correctly classified. Because of this, using accuracy as a multi-label classification metric to evaluate the performance of multi-label models is more strict, less informative, and less desirable (Helland, 2023).

3. Methodology

The methods used to address the research topics raised in this thesis are described in this chapter. To obtain the results that will be presented, this chapter aims to give a thorough description of the data collection and preprocessing, feature extraction, and model selection. To ensure reproducibility and provide the reader with a fair basis for evaluating the employment of models on text categorization tasks, this section should detail the techniques.

3.1 Data Collection

The collection of data is a crucial step that lays the foundation for later analyses and model development. It involves gathering textual data from various sources, which could include websites, databases, or specialized datasets for the specific domain of interest. It is important to collect a sufficiently diverse amount of data to capture the variability present in real-world text data (OpenAI, 2024). This thesis uses a BBC news dataset containing 2225 text data and five categories of documents (Text Document Classification Dataset, 2024).

3.2 Text preprocessing

Preprocessing methods are an essential step for text mining techniques and applications. The data and its columns need to be analyzed and inspected since it is often necessary to generate a new column combining the various features. Through the joint column, a better, more comprehensive, and more accurate analysis can take place. The three essential preprocessing steps — extraction, lowercase conversion, and StopWords removal — are covered in this study (Figure 1).

3.2.1 Extraction / Tokenization

Tokenization is the process of splitting sentences into individual words, characters, and punctuation, which are referred to as tokens. The split function uses white spaces or punctuations as dividing criteria. These generated tokens are often stored in a list afterward. In later processing phases, this step aids in removing unnecessary terms (Tabassum & Patil, 2020).

For example:

“This is an example sentence for the showcase of tokenization!”

Will be split into:

“This”, “is”, “an”, “example”, “sentence”, “for”, “the”, “showcase”, “of”, “tokenization”, “!”

3.2.2 Lowercase Conversion

Text typically consists of capital letters and abbreviations. Although this stage of text preprocessing is frequently skipped, it is one of the easiest and most successful ones. NLP is case-sensitive, meaning, it interprets 'Hello' differently than 'hello' and leads to a different outcome. In the later phases of word embedding would it create two distinct vectors, for the same words with one in capital and one in lowercase. For this reason, the best practice in text pre-processing has been to make all words lowercase (Tabassum & Patil, 2020).

3.2.3 StopWords Removal

Simple words like "the", "are", "is", "and" and so forth have no significance except in certain particular use cases. For instance, these extra words are not given any weightage in the text classification use case. The keywords that define the topics are the only ones that are extracted. Therefore, to reach the best result of algorithms these StopWords have to be found and removed from their document. It is also important to remember that in some scenarios, such as conversational models, the inclusion of specific negation words, like "No", "cannot", "wont" and "not", is crucial (Tabassum & Patil, 2020). Libraries like nltk and sklearn already offer predefined lists of StopWords which can be easily downloaded and implemented into the code.

3.3 Feature Extraction

The encoding of features in vector forms for machine comprehension is often referred to as feature extraction. After being extracted by these methods, every feature is finally represented as a vector, which is then sent to the classifier models. The most common techniques, such as Bag-of-Words (BoW), and TF-IDF will be discussed next.

3.3.1 Bag-of-Words (BoW)

A bag of words in terms of natural language processing is a collection of words based on the number of occurrences in a given text or document. It only counts the frequency of a word, regardless of its position in the text. Thus, the BoW interprets that documents or texts containing similar words share the same context. One flaw of the model is that it prioritizes words that appear more frequently, making them more significant. On the other hand, some words may occur more frequently than others but lack sufficient information to help in clustering or classification issues. Additionally, longer documents provide a greater rate than shorter ones, which reduces the accuracy of the BoW model (Tabassum & Patil, 2020).

```
Document 0: "The quick brown fox"
Document 1: "Jumped over the lazy dog"
Document 2: "The dog chased the fox"

Vocabulary:
{'the': 8, 'quick': 7, 'brown': 0, 'fox': 3, 'jumped': 4,
'over': 6, 'lazy': 5, 'dog': 2, 'chased': 1}

BoW:
    brown  chased  dog  fox  jumped  lazy  over  quick  the
0     1      0      0    1     0      0     0      1     1
1     0      0      1    0     1      1     1      0     1
2     0      1      1    1     0      0     0      0     2
```

3.3.2 Term Frequency-Inverse Document Frequency (TF-IDF)

A numerical measure called Term Frequency-Inverse Document Frequency (tf-idf) shows how significant a word is to a document within a collection. It is mostly used as a common weighting factor in text mining and information retrieval. The value of tf-idf rises in direct proportion to the frequency of a word in the corpus, but this is offset by the term's frequency in the document. This can help in managing the fact that certain words are typically used more commonly than others. Stop-word filtering using Tf-IDF is effective in a variety of subject areas, such as text classification and summarization. The model is the product of the two aforementioned statistics, termed frequency and inverse document frequency. The number of occurrences with which each term appears in each document is counted and added together to further differentiate them (Vijayarani et al., 2015). By calculating the log of the ratio of all documents to all instances of a word in a given document, it essentially scales down the less important words (Tabassum & Patil, 2020).

```
Document 0: "The quick brown fox"
Document 1: "Jumped over the lazy dog"
Document 2: "The dog chased the fox"

Vocabulary:
{'the': 8, 'quick': 7, 'brown': 0, 'fox': 3, 'jumped': 4,
'over': 6, 'lazy': 5, 'dog': 2, 'chased': 1}

  brown  chased  dog  fox   jumped  lazy   over  quick   the
0  0.58   0.00  0.00  0.44   0.00   0.00   0.00   0.58   0.35
1  0.00   0.00  0.38  0.00   0.50   0.50   0.50   0.00   0.30
2  0.00   0.53  0.40  0.40   0.00   0.00   0.00   0.00   0.63
```

4. Results

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

5. Discussion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Bibliography

A. List of Interview Partners

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi

nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

B. Code Table

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.