Real-time Carbon Neutrality Management And Optimization Using Natural Language Processing

Project Id: TMP-2022-175

Project Proposal Report

S.Vishakanan

B.Sc. (Hons) in Information Technology Specializing in Data Science

Department of Computer Science and Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

Feb 2021

Real-time Carbon Neutrality Management And Optimization Using Natural Language Processing

Project Id: TMP-2022-175

Project Proposal Report

B.Sc. (Hons) in Information Technology Specializing in Data Science

Department of Computer Science and Information Technology

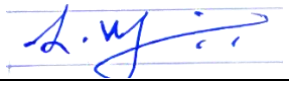
Sri Lanka Institute of Information Technology

Sri Lanka

Feb 2022

Declaration

I acknowledge that that is my very own work and this proposal does not no longer content without acknowledgment any works formerly submitted for a diploma or degree in every other college or institute of better studying and to the first-rate of our knowledge and notion it does not no longer include any material formerly posted or written through any other man or woman besides in which the acknowledgment is made with inside the text.

Name	**Student ID**	**Signature**
S.Vishakanan	IT19001562	

The supervisor/s should certify the proposal report with the following declaration.

The above candidateis carrying out research for the undergraduate Dissertation under my supervision.

---------- ----------

Signature of the supervisor Date

Abstract

Carbon dioxide has been identified as the primary cause of global climate change, which has gotten a lot of attention around the world. Among all greenhouse gases, carbon dioxide is the most prominent gas that causes carbon emissions in the environment. The entire amount of carbon dioxide (CO2) released by human activity over time is referred to as the carbon footprint. Every activity a human takes releases carbon dioxide into the atmosphere. It is critical to understand how much carbon dioxide a person emits. Vehicle emissions, energy, firewood, air conditioning, refrigeration, and other variables can all contribute to carbon emissions. All these variables can be measured by the units in charge of them. The goal of this study is to verify each factor's measurement units,convert unverified units to the actual emission factor unit. Processing text data that is a mixture of natural language and formal languages, can be measurements or units, is required for a variety of natural language techniques. This study separates unit verification and conversion into two phases and presents a text classification technique that enables a model to perform unit classification to verify the given consumption unit with the emission factor unit as well as a unit conversion sector that allows for the transformation of different class unit into the perspective emission factor units using a unit conversion factor matrix in order to calculate carbon emissions.

KEYWORDS – Carbon footprint, Unit verification, Unit conversation, NLP, Carbon emissions calculation, Text classification, Unit conversion factor matrix

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
GHG	Green house gas
CO2	Carbon dioxide
NLP	Natural language processing
DL	Deep Learning
SVM	Support Vector Machine
PCA	Principal Component Analysis
TLDDL	Transfer Learning Discriminative Dictionary Learning
NB	Naive Bayes
LR	Logistic Regression
IPCC	Intergovernmental Panel On Climate Change

1. INTRODUCTION

1.1 Carbon Footprint

Nowadays excess carbon dioxide emissions have exacerbated global warming, posing a major threat to human society's long-term viability. The question of how to reduce carbon dioxide emissions has become one that the international community is grappling with. The total quantity of greenhouse gas (GHG) emissions produced directly or indirectly by a person, corporation, event, or product is referred to as their "carbon footprint" [5]. Specifically, the total amount of carbon dioxide (CO2) emissions for which a person or business is liable called carbon emission. The majority of human activities, such as unsustainable consumption and manufacturing methods and practices, emit GHG directly or indirectly. In terms of the amount released and the overall influence on global warming, Carbon Dioxide (CO2) is the most commonplace Greenhouse Gas emitted by human activities.

Carbon dioxide emissions are produced by both natural and man-made sources. Natural sources include decomposition, ocean release, and respiration. Cement production, deforestation, and the combustion of fossil fuels including coal, oil, and natural gas are all human-made causes [6]. As a result of human activity throughout the Industrial Revolution, carbon dioxide levels in the atmosphere have risen substantially, reaching dangerous levels not seen in the last 3 million years. Despite the fact that human-caused carbon dioxide emissions are much lower than natural emissions, they have upset a natural balance that had existed for thousands of years before human intervention. This is because natural sinks take approximately the same amount of carbon dioxide from the atmosphere as natural sources produce. This had kept CO2 levels in check and within a safe range. Human-caused emissions, on the other hand, have disturbed the natural equilibrium by adding more carbon dioxide to the atmosphere while removing none.

1.2 Carbon Dioxide emission : Human Activities

Since the Industrial Revolution, human re assets of carbon dioxide emissions have increased. Human activities such as the burning of oil, coal, and gas, as well as deforestation, often result in increased carbon dioxide concentrations within the environment. Coal, natural gas, and oil burning account for the vast bulk of all human-produced carbon dioxide emissions. Wooded area removal and other agricultural growth, as well as a few industrial operations such as cement manufacture, contribute to the relaxation.

Electricity

The electricity and heat era is the economic sector that produces the most man-made carbon dioxide emissions. This business firm produced a little amount of carbon dioxide from fossil fuels. This business operation is typically reliant on coal, which has the highest carbon-in-depth of all fossil fuels and hence has a substantial global carbon footprint. In practically all advanced nations, fossil fuel burning provides the majority of energy. Depending on the power blend of the local energy supplier, the energy consumed at home and at work is likely to have a significant impact on greenhouse gas emissions.

Transportation

Human carbon dioxide emissions are mostly emitted by the transportation industry. Transporting products and people throughout the world resulted in substantial carbon dioxide emissions from fossil fuels. Transportation-related emissions have increased fast during the 1990s, more than doubling in less than a decade. Road transportation is responsible for a significant portion of the sector's carbon dioxide emissions. Automobiles, freight, and light-duty trucks are the primary sources of emissions in the transportation industry, and emissions from all three have consistently increased.

Industry

The third best supply of man-made carbon dioxide emissions is the commercial zone. Manufacturing, building, mining, and agriculture are all a part of the commercial zone. The fundamental enterprise is production, that's divided into five categories: paper, food, petroleum refineries, chemicals, and metal/mineral goods. Enormous volumes of every shape of greenhouse gas, however in particular big quantities of CO2, are produced with the aid of using production and business operations. This is because of the reality that many business centers use fossil fuels to generate the warmth and steam required at numerous levels of the producing process. For example, companies withinside the cement zone need to warmness limestone to 1450°C a good way to remodel it into cement, that's executed with the aid of using burning fossil fuels[8].

1.3 Units Classification Using NLP

Units can be varied based on the different carbon emission factors. This study focuses on unit verification, which is widely recognized as a difficult problem due to the vast range of carbon emission parameters. Text classification is the process of categorizing textual information into ordered groups[10]. It is also known as textual content tagging or textual content categorization. Text classifiers can be used to consistently identify unit classes from given texts using Natural Language Processing (NLP) and then assign a set of pre-defined tags or classes simply relying entirely on found measures[13].

1.3 Units Coversion Concept

Conversion of units is the process of converting between different units of measurement for the same amount, usually using multiplicative conversion factors. Units of measurement are required to quantify these values. There are instances when the measuring units employed do not correspond to the measurement choice and convenience, as well as the standards required for certain processes and applications. It is critical to convert such units to the point where they can be comprehended and applied correctly.

2. BACKGROUND&LITERATURE SURVEY

2.1 Background

Text classification may be used for a variety of applications in NLP, including question answering, spam detection, sentiment analysis, news categorization, user intent classification, and content moderation[10]. Manual annotation or machine labeling are both options for text classification. Automatic text categorization is becoming more significant as the amount of text data in industrial applications grows. As the identified problem of carbon emissions are risky in that they threaten the livelihood of planet, animals, humans, and ultimately, existence as we recognize it. The quantity of carbon emissions trapped in our environment reasons international warming, which reasons weather change, signs and symptoms of which encompass melting of the polar ice caps, the growing of sea levels, the disturbance of animals' herbal habitats, severe climate events, and such a lot of extra bad facet outcomes which are risky to the planet, to human and animal existence.

To limit carbon emissions, first determine how much carbon is emitted into the environment on a daily basis by a person's actions. It's difficult to quantify a person's total carbon emissions for a day since there are so many various sources that might emit carbon dioxide into the atmosphere, and the measurement results collected from the user can be in different units. The units must be classified according to the identified carbon emission sources. To categorize the various units into their perspective categories, a text classification model may be trained and applied.

A text classification technique using natural language processing will be used to classify classes of the unit provided by the user and units in the selected emission factor. It will be verified if the classes are the same using the detected classed similarity of the units. If those classes differ, a conversion matrix for distinct unit classes will be

used to identify the conversion factor for unit change, and values will be converted using the conversion factor before calculating emission.

2.2 Literature Survey

2.2.1 Research A: Natural Language Processing Techniques for Extracting and Categorizing Finding Measurements in Narrative Radiology Reports

This paper presents quantitative outcome criteria may aid in the development of a healthcare organization in which clinical procedure outcomes are repeatable and predictable. Measurements are the most common type of quantitative parameter in imaging research. They created two natural language processing engines to extract and categorize data from narrative radiology reports. In a formal evaluation, both engines performed satisfactorily. In two technical appendices, the engine's more knowledge-intensive components have been made public. The research helps to machine interpretation of radiological reports and could be used in medical data processing software.

2.2.2 Research B: Applications of Deep Learning in News Text Classification

This research take place of News Text classification approach based on the combination of deep learning (DL) algorithms. The following three innovations are included in this paper, which overcomes the difficulties listed above. The first step is to create a two-level categorization model. The first-level model is used to identify news events, while the second-level model is used to classify them. Second, unlike previous research that used a word vector to express text feature information, this paper proposes a discrete vector to express text feature information, taking into account the contribution of each word in classification and calculating the probability variance of each word to obtain the contribution of each word in classification. Finally, in the suggested model, the text characteristics are expressed using the word vector and the dispersion vector. The word vector is used to capture the semantic information between words, whereas the dispersion vector is used to show the link between words and categories. The recognition rate of the first-level model is 99.5 percent, and the accuracy rate of the second-level model is 94.82 percent, indicating that the model has considerable news event identification and classification capacity, according to the experimental comparison and analysis presented in this work. The public stop word list is utilized in text preprocessing, but the special stop word list for news events is not produced, resulting in some feature information being filtered out. The specific stop word list for news events can be established later.

2.2.3 Research C: Efficient English text classification using selected Machine Learning Techniques

This study's main purpose is to employ Weka as an experimental tool for text classification, feature selection, and performance evaluation. R, Tensor Flow, Python, or a Matlab simulation program may all be utilized to implement our strategy. Pre-processing was utilized to choose the optimal feature, such as frequency, starting letter, paragraph, question mark, and full stop, at the start. We extract text features from English language-based texts using supervised machine learning. To provide a comparison analysis, we also assessed different machine learning approaches such as NB, SVM, and LR. The simulations show that the SVM outperforms the other machine learning algorithms on the datasets we looked at. Some selected metrics, such as accuracy, recall, and F1 value, were used in the assessment and comparison. Finally, they discuss the performance of the machine learning approaches we picked. It is clear from our explanation that each machine learning classification approach has advantages and disadvantages that are dependent on the size of the datasets.

2.2.4 Research D: Sentiment Classification of News Text Data Using Intelligent Model

A transfer learning classification system for cross-domain text sentiment categorization was proposed in this work. They suggested incorporating subspace projection and transfer learning into the dictionary learning framework, inspired by the benefits of dictionary learning in knowledge reconstruction and sparse representation. They define the following discrimination information preserved term in the objective function, taking into account the within-class minimizing and between-class maximizing of sparse coding coefficients; in the meantime, they use the PCA term in the objective function to retain the discrimination knowledge. To build a relationship between multiple domains, a domain-invariant dictionary is generated in such an algorithm. The TLDDL algorithm achieves high classification performance, according to the findings of the experiments.

2.2.5 Research E: Deep Learning Based Text Classification: A Comprehensive Review

In a variety of text classification tasks, such as sentiment analysis, news categorization, question answering, and natural language inference, deep learning-based models have outperformed traditional machine learning-based techniques. They analyze the technical contributions, similarities, and strengths of more than 150 deep learning-based models for text categorization that have been built in recent years in this study. They also give a rundown of more than 40 extensively used datasets for text

categorization. Finally, they address future research prospects and present a quantitative comparison of the performance of several deep learning models on prominent benchmarks.

3. RESEARCH GAP

As per the literature review, there are several researchers have been done on the domain of text classification using NLP. Previously text classification concept was applied on different kind of topics such as movie reviews, question and answering, news text classification, spam mail detection, topic labeling, tagging content , etc. The majority of these applications have to take the unstructured text data to process, classifying that kind of text is very extremely useful for variety of such purposes. Based on the past research analysis, there were no efficient algorithms or methods developed on classifying mathematical units of measurements for any of calculation purposes. Some of the scientific report analysis resaerchers have been focused on categorizing measurement from the documents. Class comparison measures, also known as "comparison measurements," generally indicate the dimensions of a current finding but are compared qualitatively to the dimensions of the same finding on a previous. When the dimensions of a discovery on current and past are equal, it may be more artistically preferable to use a comparative measurement and other calculations[1]. The past researchers have been used to approach the text classification in three different ways as rule-based methods, machine learning based methods, hybrid methods. The rule-based methods employ handmade linguistic rules. Making a list of terms linked to a specific column and then judging the text based on the occurrences of these words is one technique to organize content. Words like "fur," "feathers," "claws," and "scales," for example, might aid a zoologist in locating animal-related writings on the internet. These methods need a great deal of subject expertise, take a long time to compile, and are difficult to scale.

To anticipate new text categories, a machine learning approach is used to train models on massive amounts of text data. The process of transforming text input into numerical data in order to train models is known as feature extraction. Two popular feature extraction methods are bag of words and n-grams. We may use a number of machine learning approaches to categorize text. The machine learning based approaches have been categorise under naive bayes classifiers, support vector machine, deep learning algorithms. By comparing with traditional machine learning based techniques, deep learning models can perform well[10]. Hybrid methods combine the two techniques mentioned above. They create a classifier that can be fine-tuned in particular instances using both rule-based and machine learning approaches.

Related-Works	Classification of mathematical units/measurements	Deep learning based approach	Comparission of units/ measurements	Convert values
Research A	✓	✗	✓	✗
Research B	✗	✓	✗	✗
Research C	✗	✓	✗	✗
Research D	✗	✗	✗	✗
Proposed System	✓	✓	✓	✓

Table 1: Comparision of Unit Classification With Former Researches of Text Classification

The aforementioned characteristics were identified as the primary research needs in previous studies and evaluations of text categorization algorithms utilizing NLP. The suggested system will fill the research gaps that have been discovered. The use of deep learning methodologies to classify mathematical units will aid in the development of new NLP concepts and techniques in the text classification arena.

4. RESEARCH PROBLEM

Carbon emission is the known problem that cause the global warming resulting climate changes[7]. Every human being should know their cumulative amount of carbon that can be emitted to the environment by his/her own day to day activities.

This is probably because of a loss of studies approximately the significance of carbon footprint measuring in the company. Every year, carbon emissions must be measured in equal kilograms of CO2 (kg CO2 e) or equal heaps of CO2 to estimate how tons carbon every character emits (ton CO2 e). This calculation will yield an annual carbon footprint figure, permitting for a higher information of carbon footprint and the method of mitigation measures[11]. "kgCO2e/(m2-a)" is the functional unit, which represents the CO2 equivalent per square meter of building area per year. This unit of computation efficiently eliminates the impact of variable building sizes and design years, ensuring that accounting results are uniform and comparable[12].

Each year, a typical passenger car emits around 4.6 metric tons of CO2. This means that the typical gasoline automobile today has a fuel economy of gallons per gallon. When a gallon of gasoline is burned, roughly 8,887 grams of CO2 are released[8][9].

The emission factor approach, commonly known as the IPCC inventory method or the emission factor method, is a popular method for assessing carbon emissions today. Carbon emission factors must be determined before carbon emission estimates for buildings can be performed[12].

The environment may be used to identify numerous carbon emission variables. The many factors may make it difficult to determine the right emission factor unit. Because a person's input may be incorrect, the overall quantity of carbon emissions may be incorrectly calculated. When utilizing a voice discussion to collect emission information, each informant has their unique manner of describing the emission unit. In terms of unit measurements, emission records come in a variety of forms. As a result, before computing a person's total carbon emissions for a day, the user's stated figures should be checked.

The validation cannot be done manually; instead, it must be verified by some form of intelligence support system. Users can select emission factors with different units, resulting in the given carbon emission unit and the values do not match the actual carbon emission factor unit, and users can select emission factors with different units, resulting in the given carbon emission unit and the values do not match the actual carbon emission factor unit. The provided unit may be converted to compute the carbon emission by determining the conversion factor.

5. OBJECTIVES

5.1. MAIN OBJECTIVES

The main objective of this research component is to design and implement a unit verificarion and conversion component to classify different units, convert them to same format. When computing carbon emissions, keeping track of units is critical. Emission factor standards provide values in a variety of units. It's crucial to make sure the emission factor and activity units are the same. This is a time-consuming task for the person in charge. As a result, the goal of my study is to determine the rate of carbon emission in a common unit that can be used to compute the rate of carbon emitted by a given element.

5.2. SPECIFIC OBJECTIVES

- Research the ways to create unit classification using natural language processing.
- Collect carbon emission units and measurements related data set.
- Identify the real units of emission factors gathered in the data set.
- Preprocess collected text dataset to train the model.
- Choose the best algorithm to classify different units.
- Implement unit conversion factor matrix to identify the conversion factor to convert the different units of measurements.

6.METHODOLOGY

6.1 System Architecture

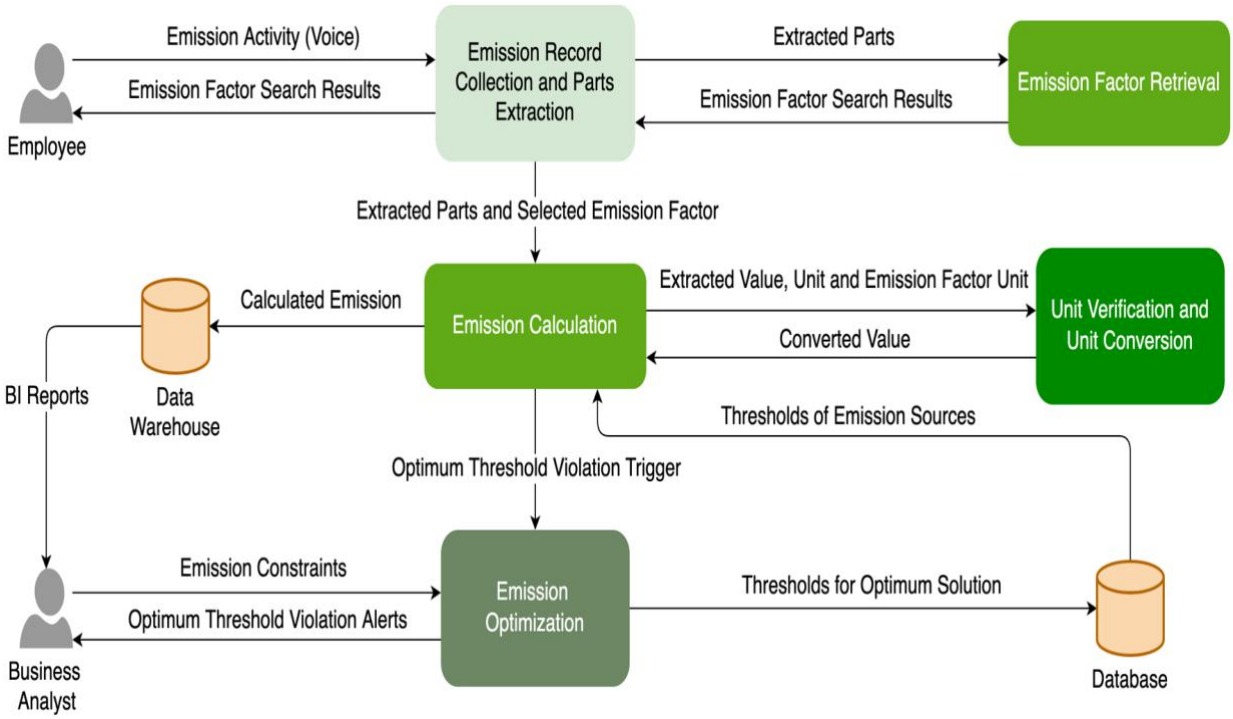


Figure 1: High level System Architecture Diagram

In figure 8.1, it illustrates the high-level system overview diagram of the Carbon emission calculation system which is proposed. There are mainly two ways a user can input the carbon emission detail for a day such as question and answering as a text or voice input. Natural language processing techniques will be used to extract necessary bits from the audio recording. Text is generated from the speech input. To extract carbon emission information, the text data is preprocessed, converted, and cleaned. The voice assistant will ask the user to explain if there is any uncertainty or missing information. It will also receive confirmation on emission parameters that have been identified as important. A factor finding module will return appropriate factors to the voice assistant for confirmation based on the emission activity sections detected by the voice assistant. Due to the technical nature of the emission factors, a natural language-

based information retrieval system will be employed to quickly locate important emission factors.

The given emission detail is subjected to unit validation and conversion in order to categorize the units of measurement and match the emission factor data provided by the user. A text classification strategy employing natural language processing will be used to classify classes of the unit given by the user and units in the specified emission factor. It will be checked if the classes are the same using the detected classed similarity of the units. If those classes differ, a conversion matrix for distinct unit classes will be used to calculate the conversion factor for unit change, and values will be converted using the conversion factor before computing emission.

The emission optimization sector gives the alerts to the user based on the maximum emission value provided by the solution, maximum thresholds are exceeded. These identified subsystems are:

1. Efficient emission factor searching.
2. Real-time collection of emission data.
3. Unit checking and conversion to match the emission factor's unit.
4. Emission optimization for the emission source constraints.

6.2 Individual Component Architecture

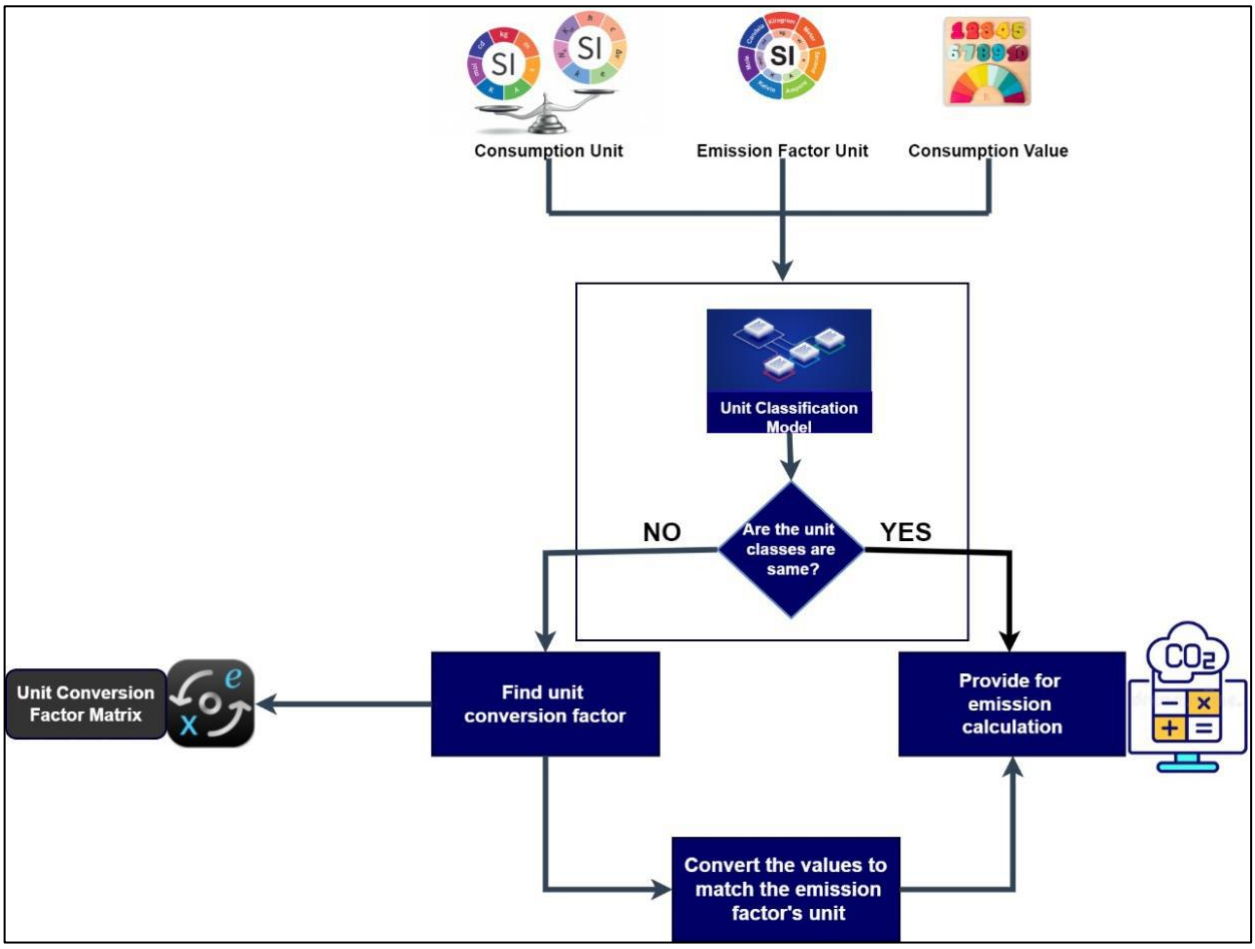


Figure 2: High level Diagram of Unit Classification And Conversion Sub System

The Individual research component which is conducted by myself includes the unit verification and conversion subsystem where the units can be verified along with the provided factor of carbon emission.The high-level architecture of the subsystem is shown in Figure 6.2. The measurements and the units of the carbon consumption is given to the subsystem with the emission factor unit which has been identified by the system. Unit classification model is used to verify if the unit classes are same or not. The consumption unit and the emission factor unit is different the unit conversion factor need to be identified from the unit conversion factor matrix and the different measurement values will be converted to the same emission factor unit. After finish the unit classification and conversion process, the output will be given for the carbon emission calculation.

The goal of this stage is to classify the different units and convert different units to the same format to calculate the carbon emission efficiently.

6.3 Implementation

The implementation process of Unit Verification Model is divided into 5 main subtasks, which are,

1. Data set collection
2. Preprocess Phase
3. Feature Extraction Phase
4. Training Phase
5. Testing Phase
6. Evaluation Phase

6.3.1 Data set collection

For the implementation of the unit verification model, we will be needing text data consist of different units and measurements of carbon emission detail. In this stage, we are focusing on collecting the datasets from our collegues using the online survey to collect various carbon emission detail. The collected data set will be having the a set of emission factors and their given measurements and units.

6.3.2 Preprocess Phase

After collecting the needed data set, to be preprocessed to filter unstructured data. Preprocessing steps is a necessary step that is focused on converting text data into a machine-readable format.

6.3.3 Feature Extraction Phase

The feature extraction phase is a very critical part of creating a model and training it. After this phase data will be ready to train a model, therefore this step is needed to be done very carefully because it will affect the accuracy of the model. Words in the text dataset consist of discrete and categorical features which must be mapped to real-valued vectors in order to be used by the algorithms. The given emission factors of the data set needed to be find the correct emission factor units. After find the real units, several techniques can be used to convert a text dataset into a vector. The vector will be consist of the emission factors and the measurements, given emission factor units , real emission factor units to train the model.

6.3.4 Model implementation phase

The model will be implemented to learn and classify the different carbon emission units from the given dataset. The different unit will be sent to the find the conversion factor from unit conversion factor matrix and the different measurements value will be converted.

6.3.4 Training Phase

The model's performance during training will decide how successfully it categorizes units in the end. When training the model, both the quality of the training data and the method used must be taken into account. Training data is divided into two categories: training and validation and testing. Algorithm-model complexity, performance, interpretability, computer resource needs, and speed are all things to consider. Due to the necessity to balance so many objectives, algorithm selection may be time-consuming and difficult.

6.3.5 Testing Phase

After the model has been trained, it must be tested using data that has never been seen before. The test results must be assessed against the performance evaluation criteria. To prevent the problem of unstable data, repeat this step by altering the training and test data during the training and testing procedure.

6.3.6 Evaluation Phase

Following the completion of the preceding four phases, the evaluation procedure begins. The major goal of this phase is to determine the trained models' accuracy and efficiency in classifying different units of the given input. I intend to analyze the system by comparing different loss and accuracy of the model, as well as collect different amount of data set to determine which part need to be enhanced and improved. Following the integration of the entire product, we want to conduct a survey for students who are chosen at random.

6.3 Gantt Chart

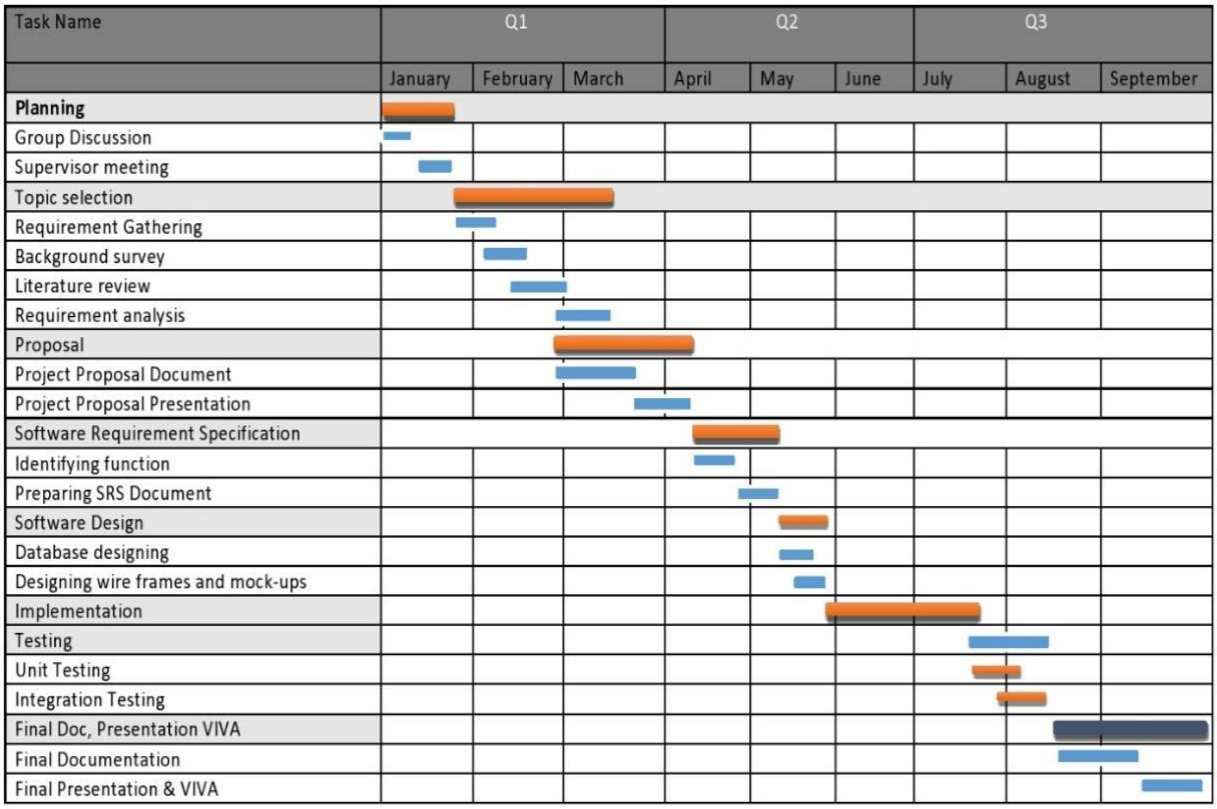


Figure 3 : Gant Chart

Commercialization

We are undertaking this study to identify industrial businesses that can use an automated method to compute each employee's carbon emissions to the environment in a timely way. Furthermore, this study focuses on how employees use carbon emission resources, and based on such usages, a minimal ideal solution for reducing carbon emissions to the environment may be supplied. In the future, this research will eventually include every single human being, allowing them to quantify their carbon emissions to the environment.

REFERENCES

[1] Sevenster M, Buurman J, Liu P, Peters JF, Chang PJ. Natural Language Processing Techniques for Extracting and Categorizing Finding Measurements in Narrative Radiology Reports. *Appl Clin Inform*. 2015;6(3):600-110. Published 2015 Sep 30.

[2] S.L., Berrahou & Buche, Patrice & Dibie-Barthélemy, Juliette & Roche, Mathieu. (2013). How to extract unit of measure in scientific documents?

[3] Hundman, K., & Mattmann, C. A. (2017). Measurement context extraction from text: Discovering opportunities and gaps in earth science. In arXiv [cs.IR].

[4] Bozkurt, S., Alkim, E., Banerjee, I., & Rubin, D. L. (2019). Automated detection of measurements and their descriptors in radiology reports using a hybrid natural language processing algorithm. Journal of Digital Imaging, 32(4), 544–553.

[5] Department for environment, food and rural affairs Nobel house 17 smith square London SW1P 3JR telephone: 020 7238 6000 website: Www.Defra.Gov.Uk. (n.d.). Gov.Uk. Retrieved February 2, 2022

[6] Chen, X., Shuai, C., Wu, Y., & Zhang, Y. (2020). Analysis on the carbon emission peaks of China's industrial, building, transport, and agricultural sectors. The Science of the Total Environment, 709(135768), 135768.

[7] Dumortier, J., Hayes, D. J., Carriquiry, M., Dong, F., Du, X., Elobeid, A., Fabiosa, J. F., & Tokgoz, S. (2011). Sensitivity of carbon emission estimates from indirect land-use change. Applied Economic Perspectives and Policy, 33(3), 428–448.

[8] Gryparis, E., Papadopoulos, P., Leligou, H. C., & Psomopoulos, C. S. (2020). Electricity demand and carbon emission in power generation under high penetration of electric vehicles. A European Union perspective. Energy Reports, 6, 475–486.

[9] Samaras, C., Tsokolis, D., Toffolo, S., Magra, G., Ntziachristos, L., & Samaras, Z. (2018). Improving fuel consumption and CO2 emissions calculations in urban areas by coupling a dynamic micro traffic model with an instantaneous emissions model. Transportation Research. Part D, Transport and Environment, 65, 772–783.

[10] Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2020). Deep learning based text classification: A comprehensive review. In *arXiv [cs.CL]*.

[11] Malek, A., & Kumarasan, K. K. (2019). Design and development of a carbon footprint calculation model for Universiti Tenaga Nasional. *International Journal of Recent Technology and Engineering*, *8*(4), 6236–6239.

[12] Shang, Mei & Geng, Haochen. (2021). A study on carbon emission calculation of residential buildings based on whole life cycle evaluation. E3S Web of Conferences. 261. 04013. 10.1051/e3sconf/202126104013.

[13] Zhang, S. (2021). Sentiment classification of news text data using intelligent model. *Frontiers in Psychology*, *12*, 758967.

[14] Luo, X. (2021). Efficient English text classification using selected Machine Learning Techniques. *Alexandria Engineering Journal*, *60*(3), 3401–3409.