Using AI for Planning Predictions – Development of a Data Enhancement Engine

Ashwath.K[1], Vikas Patel[2], Viranj Patel[3], Bhargav Dave[4]

[1]Post Graduate Student in Construction Engineering and Management, Faculty of Technology, CEPT University, Ahmedabad, Gujarat

[2]Technical Support Specialist, VisiLean, Ahmedabad, Gujarat

[3]Product owner, VisiLean, Ahmedabad, Gujarat

[4]CEO, VisiLean, Ahmedabad, Gujarat

ashwath2917@gmail.com, vikas.patel@visilean.com, viranj.patel@visilean.com, bhargav@visilean.com

Abstract

The construction industry is prone to cost overruns, time overruns, and quality issues. These are all caused by several factors, among which improper planning resulting from unreliable assumptions and unpredictable risks plays a significant role. This could be overcome if the planning-related issues are well predicted, and the risks are accounted for appropriately. Prediction of the various aspects of planning could be enabled by identifying the trend in past data and making it useful for current and future planning. Also, the data being semantic is significant to perform decisive predictions. This could be done by leveraging growing technology like Machine Learning, and Natural Language Processing which are all a part of Artificial Intelligence. Hence, as a starting point for building a robust prediction engine, this research focuses on building a data enhancement engine that would link the plan and model automatically, enabling seamless correlation of the data in the two silos and making it semantic. An engine has been developed as part of this research, and it performs well with respect to the existing data set. Several suggestions to improve the engine have been given as part of the future scope.

Keywords

Artificial Intelligence, Machine Learning, Planning predictions, BIM, Data Enhancement, Sustainability.

1 Introduction

The construction industry is a sector that has the capacity to massively impact the economic and social development of any country. It has grown tremendously in the past years and is also expected to grow steadily in the upcoming years. It determines a country's technological and technical advancement and regulates the growth of the country's infrastructural development, often directing to its advancement with respect to its sustainability assurance [1]. But the challenges related to cost, time, and quality are predominant in the construction industry [2] and are caused by improper planning, which is a significant determinant of the success of any construction [3],[4]. Planning includes proactive identification and resolution of day-to-day issues, make-ready process, linking short-term and long-term planning, proper duration estimation, proper resource planning, prediction of quality issues and establishing control over them, proper design management, etc. These factors vary based on the scale of a project, its nature, region, etc. These factors are highly interdependent, making it difficult for planners to identify potential risks and make reliable assumptions during the planning process. This risk can be averted if the planners are familiar with these aspects of planning by exposing themselves to the past information corresponding to these factors [5]. Planners need to understand them by correlating the data corresponding to the various factors involved in planning.

The data we are talking about here is not obtained from a single source. It is extracted from several components of the construction industry like the material and equipment sector, consultancy, site execution, investment and developer database, and databases linked to technology like BIM, etc as shown in Figure 1. However, reviewing the data in an isolated manner would not give a broad picture of any situation leading to improper assumptions. These data do not become information unless correlated, streamlined, and made reliable for the planned project. The more reliable the information, the higher would be the accuracy of predictions leading to proper planning.

It is nearly impossible for planners and construction managers to manually assess the enormous reserves of past data and draw parallels between them and the current projects and remember the inferences from such mass data. Unless there is a mechanism to use the past data effectively, the decision taken by the manager based on their experience would be subjective [6].

Artificial Intelligence (AI) enables Machine Learning (ML) and Natural Language Processing (NLP) which could be leveraged to effectively collect, organize, and derive effective inferences from existing and past data. [7]. They are capable enough to establish hidden patterns among the data [6] and effectively pull out the interdependencies among them. This would help deriving insight from the existing data thereby converting data into information.

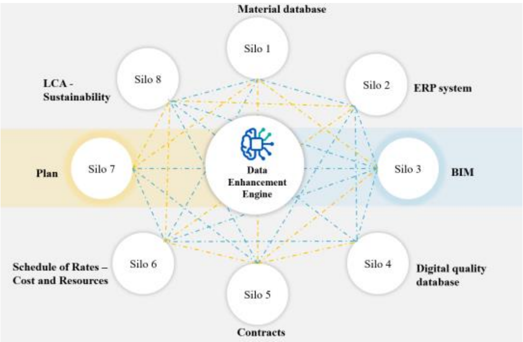


Figure 1: Data sources and corelation

Among the several sources of data mentioned, this research focusses on the plan (i.e., schedule integrated with resource information) and data-rich model (i.e., BIM-based 3D model) which are predominant in the industry. The engine to be developed would effectively link the two silos automatically, creating vast reserves of interlinked data, which would be the base for performing predictions. AI, ML, and NLP-based technologies have been explored concerning the data available, and solutions have been arrived at to perform automatic linkage of the plan and model. The outcome would be a novice engine that could predict the linkage between the plan and model to a certain extent.

1.1 Complexity in existing data extraction

Vast contextual data is available in the BIM models and project plans. This data being in two silos would be less valuable and would not open the opportunity to several use cases like 4D, 5D, planning predictions, etc. unless the data from the silos are linked and correlated with each other. The linking of the model elements and the tasks is not done precisely because of the difficulties in the existing linkage methods. The existing methods to link tasks are as follows:

- Manual linking by selecting the task and the model element
- Linking based on any common ID defined in both the task and the model element

The first method makes the planning process difficult for the last planners. It would even delay the planning process. The second method makes the modeler's work difficult since grouping and labeling the model elements with an ID is time-consuming. Once these difficulties are overcome, the resulting data which could be extracted would be richer in context.

1.2 Aim and Objective

The aim is to develop an AI-driven engine to link the schedule activities and model elements automatically.

The objectives to be achieved to achieve the main goal involve the following.

- Develop a suitable BIM model and relevant plan/schedule for the selected projects
- Leverage the model, schedule, and the AI-ML technology to prepare a proof of concept
- Validate the POC based on relevant metrics and improve the engine accordingly

1.3 Scope

For this research purpose, the scope for data includes three projects. 2nd and 3rd projects are for validating the effectiveness of the POC.

- Project 1: This is for developing the POC. This is a multistorey commercial cum residential complex
- Project 2: This is a multistorey commercial cum building
- Project 3: This is a museum building with RCC and steel structural elements.

2 Research methodology

This research involves active participation in an organization development of an engine that would solve the mentioned problems in the industry. The objectives to develop this engine have been laid down; the engine would be developed, experimented to provide a solution for certain instances of the problem, evaluated based on relevant metrics, and the development would be iterated for perfection ending up in generalizing the perfected engine to the industry. Based on this, the design science approach seems to be the best suitable framework to base the research rather than the other existing research approaches. [8] have given six steps that make the constructive research approach and the research has been carried out until the fifth step i.e., development, demonstration, and evaluation of the engine. Further, several optimizations to the solution have been suggested, which would be carried out through iterations as a part of future scope which is the sixth step of the design science approach.

A digital platform that enables the linking of a plan and a BIM model is necessary for the research. Several platforms like Navisworks, Synchro, VisiLean etc., are available to do the same. Apart from conventional information included in a plan like schedule, cost, and resource data, more Last Planner System (LPS) based information like constraints, reasons for variance, etc., could be added in VisiLean. This makes the data more contextual and useful for predicting planning issues. Hence, VisiLean has been chosen to be the platform for linking the plan and model and retrieving the necessary data related to the linked plan and model for the research.

3 Data Collection and Processing

3.1 Collection and Preparation

Data collection and preparation involves preparing the model data, relevant schedule data, and the linkage data necessary to teach the machine for automatic linkage. These are as explained below.

3.1.1 Preparation of model, schedule and extraction of data

Model corresponding to Project 1 has been obtained from the organization. Model data are obtained through the VisiLean system by using a certain API program. Data regarding both the structural and architectural models have been obtained in such a way.

The schedule corresponding to the model has been prepared in Primavera. The schedule is then imported into VisiLean and the entire schedule related information can be extracted from the software.

3.1.2 Linking and extracting linkage data

The linkage of the activity and model elements has been done manually in VisiLean. The columns, beams, and floors from the structural model and the architectural model's masonry walls were linked to the respective activities. There are no separate elements for shuttering or reinforcement in the model, and hence the activities related to shuttering and reinforcement are also linked to concrete elements. Staircase, Lift and Finishing work related activities are not linked because of absence of relevant elements in the model.

After the linkage, two files corresponding to the linkage i.e., one for the structural model and other for the architectural model have been obtained. These files mainly consist of the information regarding the activity and element which are linked together. It doesn’t have the information related to those activities and elements which are not linked.

3.1.3 Data Summary:

Model, schedule, and linkage data have been obtained by the end of the data collection and preparation process.

3.2 Data Processing

The data obtained cannot be used directly for the ML algorithm. The data should be processed and brought to the necessary format to proceed further. The various processing steps followed are as explained further.

3.2.1 Dimensionality reduction

The reduction of features is referred to as dimensionality reduction in ML terms. Features mean the columns representing the several model, schedule, and linkage parameters. Dimensionality reduction can be done using ML-related coding, but it has been done manually here. The data available is huge enough and misleading because of the presence of unwanted and useless data. Using Power BI, each column has been analysed, empty columns and columns with very minimal and repetitive data have been removed before combining the data.

The data corresponding to string datatype i.e., only text is considered for the initial trials. The data corresponding to several other datatypes would add complexity to the algorithms to be applied. Hence, the parameters corresponding to the float data type have been eliminated for the initial trials of algorithm and the corresponding pre-processing. Certain parameters would be not necessary, and some would be repetitive. For example: Parameters like ‘Omni class title’ in the BIM model convey the same information as the parameter ‘Name’ which tells name of the element. Hence, such repeated info conveying parameters should be removed to reduce the noise in the data.

3.2.2 Combining data

Currently there are several files corresponding to schedule, model and linkage. These files have to be combined into a single file based on some common parameters in these files. The base for combining these files will be the linkage file data. It has the GUID of the activities and the BIM Model IDs of the elements which are linked together and this can be done using PowerBI.

3.2.3 Labelling data

The linkage data contains only the linked activities and model elements, and hence the label corresponding to linkage would be ‘Yes’ to all the rows. Whereas false data should also be present to teach the machine. False data means the data corresponding to the activities and elements which do not link. Hence, the data corresponding to ‘No’ linkage must be extracted.

3.2.4 Dimensionality reduction

3.2.5 Pre-processing of text

Since only text data is considered for the initial trials, text pre-processing is an important step before applying any ML-based algorithm. The parameters mentioned in Table 1 represent the data considered for the initial trials. The presence of text data necessitates the use of NLP along with ML-based algorithms. Several NLP based construction sector related projects were analyzed with respect to the methodology which has been followed and the pre-processing techniques followed in those projects.

Based on the review of the NLP based projects, several pre-processing techniques have been identified, which are essential to make the text suitable enough for the machine to analyze and run any ML algorithm on them [9],[10],[11],[12]. The pre-processing technique carried out are like punctuation removal, filling of blank cells, lowercasing, removal of stop words, tokenizing, lemmatization and vectorization.

3.2.6 Vectorization with Word2Vec and DocVec

Word2Vec is a method used to represent words as vectors, i.e., numbers in an N-dimensional vector space, i.e., it helps to obtain contextual word embeddings. The words would be placed in the vector space based on their combination and frequency of occurrence. This method indirectly considers the semantic relationship between the words since the combination of their occurrence plays a vital role in vectorization and not only the frequency.

The process involves feeding the right combination of words into the Word2Vec algorithm. After feeding them and running the model, a 2D graph could be generated to visualize the placement of words as shown in Figure 2.

The main limitation of using the Word2Vec algorithm alone is that only the individual words could only be vectorized and not sentences. To compare the plan and model data, we have several sentences that are to be compared and not only individual words. DocVec algorithm has been found to be useful to obtain the sentence vectors. Hence, the DocVec model combined with the Word2Vec model seemed to give the required results.

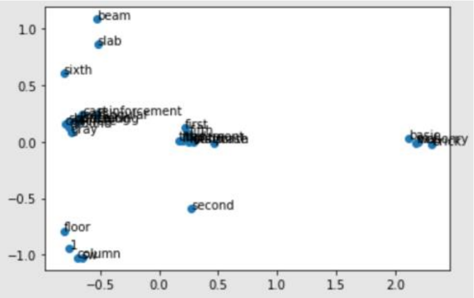


Figure 2: 2D graph output of Word2Vec model

4 Development of POC

The aim of this study is to develop an engine that would automatically link the model elements to the schedule activities to reduce the massive consumption of human resources and time required to perform the linkage process. This issue is also supported by [13] and they have devised a method to do the same. They have proposed a mapping solution with a learning loop that can adapt itself to the different activities and element categories of projects of varying nature. The mapping has been done based on three attributes namely Building, Level and Discipline.

There exist numerous other parameters in the plan obtained from VisiLean and the model. Several of these parameters would also be relevant to the linkage, which can be added to the mapping process. Although several parameters have been eliminated to reduce the complexity of the initial trials of pre-processing and application of algorithms, they can be added on further after a preliminary engine is made, which shows promise for automatic linkage. This preliminary engine is termed the 'Proof of Concept', and it is one of the main objectives of this research.

4.1 Layers to be scrutinized for mapping

From the plan, we can get the task's name, location data, and even the trade-related details. Unlike the task information, model elements have numerous parameters. All these parameters would contribute to the linkage in one way.

With respect to the data, we have for the initial trials of pre-processing and application of algorithms, from the information corresponding to plan, the task name, trade and location details can be identified. To associate these details to the model, we need similar details from the model information which we have. From the model, the element name, location details and material details could be used to relate to the plan information.

Table 1: Layers considered

Parameters	Layer
Task name and Element name	Name
Task location and Element location	Location
Task name and Element material	Trade

4.2 Application of algorithms and evaluation

A combination of the layers mentioned in Table 1 have been experimented with several algorithms to achieve the best performance.

The study is focused on supervised Machine Learning.

Based on the data and the kind of prediction we have to do, algorithms that can do effective classification are preferred to those which do regression.

Several algorithms have been used based on the problem to be addressed. The most recurrently used algorithms are Decision tree, Support Vector Machine, and Naïve Bayesian. But there is no clear distinction on which algorithm is best because the performance depends on the kind of data and purpose of prediction. Based on interaction with certain AI experts, three algorithms, namely Random Forest Classifier (RFC), Support Vector Machine (SVM), and XG Boost, were considered for the initial trials. RFC and SVM are single algorithms, whereas XG Boost is a decision tree-based ensemble algorithm.

Before the data is passed on to any algorithm, the pre-processing steps mentioned earlier must be performed. After the pre-processing is done, the similarity between the sentence vectors obtained from the Word2Vec and DocVec model will be identified. These similarities would be passed as the input to the algorithms and the 'Yes' or 'No' linkage data would be given as the output to these algorithms. Based on the input and output, the algorithms would be trained.

4.3 Combination of layers and tests

Five different combinations of the layers have been tried out to get the optimum result. This can be seen in Table 2.

Table 2: Trials and relevant details

Trial	No of layers	Layers used	Remarks
1	2	Name, Trade	-
2	2	Name, Location	-
3	3	Name, Location , Trade	-
4	3	Name, Location , Trade	In this case, the unwanted words like 'level', 'floor', 'pour' etc., were removed from location data.
5	3	Name, Location , Trade	In this case, after removing unwanted words, the locations were categorized into numbers. Ex: 'First'=1.

While applying each algorithm, the training data is split into 70% and 30% among which the 70% portion would be used for training and the remaining 30% would be used by the algorithm to test itself. This testing will be termed 'self-test' from now on in this report. Apart from this testing, five different combinations of plan parameters have been provided (combinations of words which are not a part of the self-test data set – different words and different combinations), and the predictions have been evaluated manually. This testing will be termed 'external-test' from now on in this report. The trials performed with the varying combination of the layers and the performance evaluation of the algorithms for the self-test and external-test have been discussed in detail in the following sections.

4.4 Evaluation of the tests

The performance measurement indices corresponding to the confusion matrix are predominantly being used to evaluate ML-based algorithms. Such indices referred from the literature such as Precision (P), Recall (R), F1 score (F1) and Accuracy (A) have been considered for evaluation. Higher Accuracy, Recall and Precision are preferred for a better performance.

5 Results and Discussions

The values of the indices mentioned earlier have been compared among the several trials. All these analyses correspond to the self-test performed by the algorithms based on a part of the training data set.

Parallel to the self-test performed, external testing has been carried out by providing five sets of plan data, and the predicted model data has been manually analyzed for performance. Comparing the results of the self-test and external testing results, the combination of layers was experimented with until the most optimum combination was achieved.

5.1 Performance evaluation

The legend in Figure 3 conveys the information regarding the trial, the combination of layers used and the algorithms having high performance. In some trials there are distinct difference between the performance of the algorithms and in those cases the respective single algorithm is mentioned. In the cases where the performance of algorithms varies minorly (i.e., + or – 1 or 2), multiple algorithms have been taken into consideration for high performance, and the highest value under each performance index is taken for the graph.

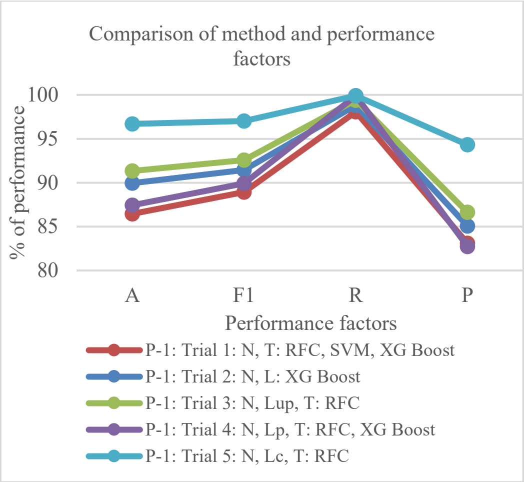
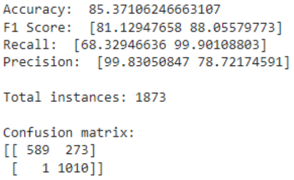


Figure 3: Trial 1-5 self-test performance indices

5.1.1 Performance of algorithms with respect to self-test as per confusion matrix related indices

In Trial-1, all the three algorithms have similar performance. In Trial-2, the performance of XG Boost is distinctly higher than the other two, and hence it is considered for high performance. In the case of Trial-3 and Trial-5, RFC shows the best performance, followed very closely by XG Boost, and the SVM shows lower performance compared to them both. In the case of Trial-3, RFC performance is distinctly higher than others. In Trial-4, the performance of all the algorithms varies minutely from each other, whereas with respect to 'Recall', XG Boost performs distinctly better than the other three, and hence we could say XG Boost performed better overall in Trial-4.



```
Accuracy:  85.37106246663107
F1 Score:  [81.12947658 88.05579773]
Recall:  [68.32946636 99.90108803]
Precision:  [99.83050847 78.72174591]

Total instances: 1873

Confusion matrix:
[[ 589  273]
 [   1 1010]]
```

Figure 4: Confusion matrix and relevant performance metrics corresponding to RFC algorithm in Trial 1

5.1.2 Evaluation of Trial

From Figure 3 we can infer that the performance increases gradually while the layers are being added one by one. All the performance factors of Trial-3 > Trial-2 > Trial-1. But, if we compare this to the performance in case of external tests, the location category has not been predicted by any of the algorithms in any of the Trials. But the prediction of location is a must, and hence the location parameter must be optimized in such a way to get good results. The reason behind the location being not predicted might be the availability of very less vocabulary with respect to the location data and the confusion created by the unnecessary words like 'floor', 'level' etc., in the location data.

Hence, Trial-4 is tried out with processed location parameters free from words which do not add much value for linking purpose. This trail will help us to understand whether removing such unwanted words increase the performance. Figure 3 shows a decrease in performance with respect to all the performance indices when compared with the first three Trials. But when we compare the previous trials with this trial, some improvement has been seen with respect to location predictions. But still, the location predictions are not satisfactory, and hence further optimization has to be done in case of the method used.

Since the lesser vocabulary in location data makes it difficult for the engine to make a clear distinction between the several locations, the available locations can be categorized into numerical values and then taught to the engine, which will help to make a clear differentiation between the locations. This categorization has been done to location data in the case of Trial-5 combined with the similarity method for name and trade-related data.

From Figure 3 we can infer that the performance of Trial-5 is much better than the first 4 trials in the case of all performance measurement indices, and this high performance corresponds to the RFC algorithm. XG Boost is the next high performing, whereas SVM is low performing compared to these algorithms.

5.1.3 Performance measurement indices

With respect to all the indices, Trial-5 scores the highest. In the case of Recall, although all the trials performed well, the Precision of Trial-5 is distinctly higher than the others. Having 100% Recall value and 94% Precision value in Trial-5, we can say that the engine would seldom predict a 'Yes' linkage as 'No' whereas it may predict a 'No' linkage as 'Yes'. But there are high chances for the engine corresponding to the other trials to predict the 'No' linkage as 'Yes', which would not be desirable.

5.1.4 Outcome

By comparing the performance of the best performing algorithms among the different trials, Trial-5 is found to be better. Name and Trade are predicted to a good extent but none of the algorithms perfectly achieved the prediction of location for all the external test instances.

The outcome of Trial-5, RFC and Brick Work activity in Second level has been highlighted in model and is as shown in Figure 5. This outcome corresponds to the results from the prediction. The rightly predicted element ID's have been obtained and used for highlighting in the model.

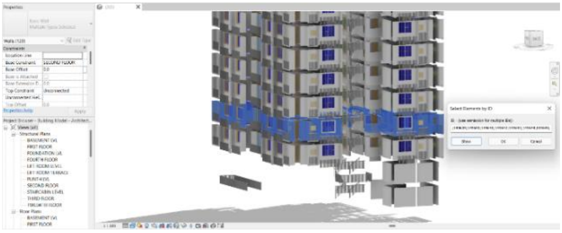


Figure 5: Right wall elements highlighted in model

We can see that all the predicted elements are correct, i.e., all are wall elements corresponding to brick material and second level. The bad performance with respect to location prediction could be oriented to the following reason. For elements like columns and walls, the location provided by the planner would match with the base levels of the model elements whereas in case of slabs and beams the locations in model would be one level ahead of the location mentioned by the planner. This is because the beam or slab of a particular floor would be placed at the base of the next floor.

5.2 Testing of POC with additional project, its results and evaluation

The second and third project data have been obtained, processed and added to the existing project data and then the machine has been taught and tested similar to the previous case. Layers corresponding to Trial-5 are considered for the combination of 1st and 2nd project as Trial-5 gave the best results in the previous case whereas layers corresponding to Trial-4 are considered for the combination of 1st, 2nd and 3rd project as the location data in the 3rd project is more specific to the nature of the project and cannot be categorised. Both self-test and external test with new plan data as input have been performed and the results have been compared to the previous case results.

The self-test results were good and similar to the previous case as shown in Figure 6 but the external test results were not satisfying and poorer than the previous case. Several instances which were predicted rightly in the previous case are not predicted properly in the case of the combined project data. In the case of the external tests run with respect to single project data, the predictions of Name and Trade were right for even new activities. Several reasons for the non-satisfactory performance are like

- The vast transformation in the Word2Vec model because of the addition of several new vocabulary and its implication on the existing combination of words.
- Insufficiency and inconsistency in data points, i.e., in general, the data points used for training are very less for a NLP-based project and comparatively, Project-1 has more data points with respect to 'Yes' and 'No' linkage compared to Project-2.

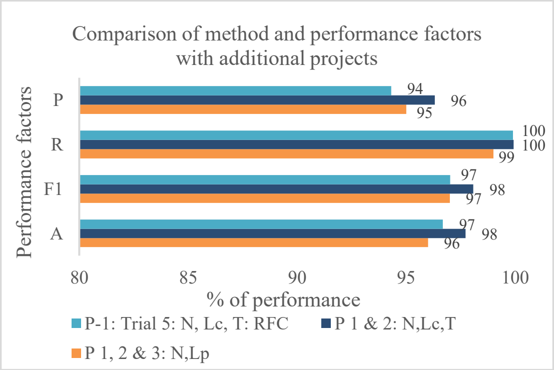


Figure 6: Self-test results of additional projects

5.3 Limitations of the engine

Several limitations are associated with the usage of AI based techniques which is applicable to this engine [7]. They are as follows.

- Biased outputs based on nature of data fed
- Uncertainty in the functionality since past data would not be available to train the system for all possible situations. This scarce nature of data availability limits the engine's scalability
- Difficulty in finding the root cause for any errors in the output since huge reserves of data is involved

5.4 Future scope: Learning loop and extended use cases

The engine can be perfected by providing it with ample data and by enabling self-learning. As the users plan through VisiLean, even though if the machine predicts wrongly, they could link the right task and element from which the engine would learn the right path of linkage and such continued learning would result eventually in the perfection of the engine. This method of continuous improvement through dynamic interaction of the planner with the system during the planning process would address the limitation of data scarcity, scalability, and in-built data bias to a considerable extent. This is a combination of edge computing and AI which is in line with the principle of Edge AI.

Once the engine is perfected, it poses 'n' number of use cases. It can be used to perfect the planning process by predicting the uncertainties with respect to duration, quality etc., with more accuracy, it can also predict sustainability parameters by accessing the existing sustainability databases as represented in Figure 1 and also by making use of the learnings from the user input sustainability factors etc.

6 Conclusion

The engine which has been developed shows good performance concerning the self-tests performed on a portion of the training data. The external test performance is comparatively not satisfactory.

It has been identified that pre-processing of data is a vital process in developing the engine. The engine developed can predict the element and trade, i.e., material to a reasonable extent but the prediction of location is not satisfactory. Main reasons identified for unsatisfactory performance are like insufficient relevant sample size, insufficient labelled data, inconsistency in the vocabulary used across the projects considered etc. Several limitations have been identified at several phases of the research. The areas to be mainly focused on are such as enlargement of sample size, effective pre-processing and the method to assess the similarity. These can be taken as a part of future scope.

If the limitations are overcome, and the engine is optimized it would provide huge reserves of contextual data essential for predictions. This engine could be used to collate and corelate data from several databases such as contract documents, standard manuals, quality related digital databases, sustainability related databases etc. This would be helpful for the various stakeholders in the construction industry to plan effectively to overcome the challenges thereby reducing the wastage with respect to materials, process, time, cost etc. and also to dynamically monitor the project sustainability. Thus, the proposed AI based engine could automate the data enrichment thereby making construction sustainable.

References

[1] Chaudhery, Mustansar Hussain, Mosae Selvakumar Paulraj, S. N. (2022). Source Reduction and Waste Minimization. Elsevier. https://doi.org/https://doi.org/10.1016/C2020-0-01110-2

[2] Ibrahim, A. R. Bin, Roy, M. H., Ahmed, Z. U., & Imtiaz, G. (2010). Analyzing the dynamics of the global construction industry: Past, present and future. Benchmarking, 17(2), 232–252. https://doi.org/10.1108/14635771011036320

[3] Hamzah, N., Khoiry, M. A., Arshad, I., Tawil, N. M., & Che Ani, A. I. (2011). Cause of construction delay - Theoretical framework. Procedia Engineering, 20(Kpkt 2010), 490–495. https://doi.org/10.1016/j.proeng.2011.11.192

[4] Johnson, R. M., & Babu, R. I. I. (2020). Time and cost overruns in the UAE construction industry: a critical analysis. International Journal of Construction Management, 20(5), 402–411. https://doi.org/10.1080/15623599.2018.1484864

[5] Asadi, A., Alsubaey, M., & Makatsoris, C. (2015). A machine learning approach for predicting delays in construction logistics. International Journal of Advanced Logistics, 4(2), 115–130. https://doi.org/10.1080/2287108x.2015.1059920

[6] Ayhan, M., Dikmen, I., & Talat Birgonul, M. (2021). Predicting the Occurrence of Construction Disputes Using Machine Learning Techniques. Journal of Construction Engineering and Management, 147(4), 04021022. https://doi.org/10.1061/(asce)co.1943-7862.0002027

[7] Harvard Business Review. (2019). Artifical Intelligence: Insights you need from Harvard Business Review. Harvard Business Review Press.

[8] Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

[9] Hassan, F. ul, & Le, T. (2020). Automated Requirements Identification from Construction Contract Documents Using Natural Language Processing. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 12(2), 04520009. https://doi.org/10.1061/(asce)la.1943-4170.0000379

[10] Jallan, Y., Brogan, E., Ashuri, B., & Clevenger, C. M. (2019). Application of Natural Language Processing and Text Mining to Identify Patterns in Construction-Defect Litigation Cases. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 11(4), 04519024. https://doi.org/10.1061/(asce)la.1943-4170.0000308

[11] Jung, N., & Lee, G. (2019). Automated classification of building information modeling (BIM) case studies by BIM use based on natural language processing (NLP) and unsupervised learning. Advanced Engineering Informatics, 41(September 2018). https://doi.org/10.1016/j.aei.2019.04.007

[12] Moon, S. M., Lee, G., Chi, S., & Oh, H. (2019). Automatic Review of Construction Specifications Using Natural Language Processing. Computing in Civil Engineering, 105–113. http://toc.proceedings.com/49478webtoc.pdf

[13] Vaidyanathan, K., Raphael, B., Kumar, A., S, M., & Patil, Y. (2021). Seamless integration of site data for planning and monitoring of construction projects.pdf. Proceedings of the Indian Lean Construction Conference - ILCC 2021.