A Sparse-Attention Framework for High-Dimensional Environmental Cost Prediction

MSc Research Project
Data Analytics

Tejaswini Matta
Student ID: 23395494

School of Computing
National College of Ireland

Supervisor: Dr. Furqan Rustam

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Tejaswini Matta ………………………………………………………………………………………
Student ID:	23395494 ……………………………………………………………………………………….……
Programme:	MSc Data Analytics …………………………………………………… **Year:** 2025 …………………………..
Module:	Research and Practicum ……………………………………………………………………………………….………
Supervisor:	Dr. Furqan Rustam ……………………………………………………………………………………….………
Submission Due Date:	28-Jan-2026 ……………………………………………………………………………………….………
Project Title:	A Sparse-Attention Framework for High-Dimensional Environmental Cost Prediction ……………………………………………………………………………………….………
Word Count:	7650 …………………………………… **Page Count** 20 …………………………………………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Tejaswini Matta ………………………………………………………………………………………………

Date: 28-01-2026 ………………………………………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

A Sparse-Attention Framework for High-Dimensional Environmental Cost Prediction

Tejaswini Matta
23395494

Abstract

Environmental sustainability assessment increasingly depends on data-driven models that can analyse complex, high-dimensional environmental indicators for accurate environmental cost prediction. However, capturing nonlinear relationships across heterogeneous sustainability features remains challenging, particularly for large-scale corporate reporting and policy-facing decision-making. Traditional analytical and statistical approaches often fail to represent complex interactions in such data and tend to underperform when variables are numerous, correlated, and nonlinearly related. Moreover, many conventional machine learning methods do not incorporate attention mechanisms to focus on the most informative features, while transformer-based models support attention-driven learning but are often computationally intensive, limiting their suitability in resource-constrained environments. To address these limitations, this study leverages a Sparse-Attention Tab Net model that combines attention-based feature selection with improved computational efficiency for environmental cost estimation. A unified evaluation framework is developed to compare classical machine learning baselines, recurrent deep learning models (LSTM and GRU), and Tab Net using consistent preprocessing, train–test splitting, and regression metrics (R^2, MSE, RMSE). Tab Net applies sparse attention to perform adaptive feature selection and stable learning on tabular environmental data, while also enabling clearer identification of influential predictors through attentive feature masks. Experimental results show that deep learning models outperform traditional machine learning baselines, with GRU and LSTM achieving strong predictive performance. The Sparse-Attention Tab Net model delivers the best overall results, achieving an R^2 of 0.98 and the lowest MSE and RMSE among all evaluated models. These findings indicate that sparsity-driven attention can improve both accuracy and feature-level clarity for robust environmental cost prediction in sustainability assessment.

1 Introduction

Sustainability of the environment has emerged as a major concern in the world because with the growing level of industrialization, there is a consequent rise in emission, waste materials and strain on the available resources. To illustrate, the CO2 generated worldwide through energy consumption in the year 2022 was a record high of 37.8 Gt, which shows that the problem is massive, and intensive monitoring and prediction instruments are required (Crippa et al, 2022). Meanwhile, the world produces approximately 2.01 billion tons of municipal solid waste annually, and this is expected to increase to 3.40 billion tons by 2050 with a substantial portion of it still being mismanaged as a result of unsafe disposal methods. The trends render the proper environmental cost estimation a core requirement of organizations and governments to facilitate compliance, investment planning, and long-term sustainability decision-making (Farooq and Khan, 2025; Ibn-Mohammed et al., 2023).

Historically, the approaches to environmental impact assessment and cost estimation have been based on the application of Life Cycle Assessment (LCA), and classical statistical/time-series modelling. LCA has been extensively applied to assessing environmental impacts in a product or service life cycle and is standardized in international guidelines like ISO 14044, which provides guidelines on the definition of goals and scope, inventory analysis, impact assessment, and interpretation. Classical methods, including regression-based models and autoregressive time-series models (e.g. ARIMA-type models) have also been widely applied in forecasting-oriented areas, but generally assume simpler functional form and may have difficulties in high-nonlinearity and when the number of variables to be predicted is large and non-independent. These classic methods can become less predictive as environmental datasets contain more and more heterogeneous indicators, missing values, and multifaceted cross-feature dependencies, or have to be heavily engineered by hand.

Machine learning (ML) and deep learning (DL) approaches help address these limitations by learning nonlinear patterns directly from data and scaling to higher-dimensional feature spaces. Prior studies show that ML can be effective but often depends strongly on feature design and data quality (Sumon et al., 2024), while DL architectures such as LSTM and GRU can better model complex patterns and interactions (Wang et al., 2025). However, many ML/DL models still face practical challenges in sustainability applications, including limited feature clarity (difficulty explaining why a prediction was made), sensitivity to heterogeneous tabular inputs, and computational overhead. These gaps motivate the need for models that can provide both high predictive accuracy and more transparent feature-focused learning behaviour for environmental cost prediction.

Research Question

To what extent can a Sparse-Attention-Enhanced Tab Net architecture improve the accuracy of environmental cost prediction compared to existing machine learning and deep learning models?

Research Objectives

1. Develop a unified environmental cost prediction pipeline using consistent preprocessing (cleaning, encoding, imputation, scaling) and an 80/20 train–test split for fair model comparison.
2. Implement and benchmark multiple models (classical ML, LSTM/GRU deep learning, and sparse-attention Tab Net) using standard evaluation metrics (R^2, MSE, RMSE).

Key Contributions

1. Computationally efficient alternative to dense transformers: Demonstrates that Sparse-Attention Tab Net can achieve transformer-level (or better) performance on high-dimensional tabular environmental data with lower computational overhead than dense self-attention architectures.
2. Attention-based learning for tabular sustainability indicators: Leverages sparse-attention masks to selectively focus on the most relevant environmental features during training, improving learning stability on heterogeneous, correlated variables.
3. Unified and fair benchmarking framework: Implements a consistent end-to-end pipeline (cleaning, encoding, imputation, scaling, 80/20 split, and common metrics: R^2, MSE, RMSE) to compare ML baselines, recurrent DL (LSTM/GRU), and attention-based tabular models.
4. Empirical performance evidence for environmental cost prediction: Shows that TabNet achieves the best overall predictive results ($R^2 = 0.98$ with the lowest MSE/RMSE), supporting its suitability for robust environmental cost estimation.

2 Related Work

The fast development of machine learning and deep learning has greatly altered the environmental sustainability research by allowing the study of complex, large and dynamic environmental data. These technologies are useful in climate modelling, green innovation, environmental monitoring and corporate sustainability assessment providing predictive services that are far beyond the capabilities of traditional methods of analysis. In spite of these developments, there are a few methodological and indicating the necessity of more transparent, efficient, and robust models that will be able to work with heterogeneous environmental data.

2.1 Machine Learning Approaches in Environmental Sustainability

The use of machine learning (ML) in environmental sustainability studies has emerged as the fundamental instrument of the field due to the ability to handle large and intricate datasets and the capacity to model nonlinear relationships which the conventional statistical methods may not capture. Farooq and Khan (2025) explain that ML enhances prediction of climate and sustainability by determining the complicated relationships between high-dimensional indicators of the environment, aiding in the enhancement of forecasting and decision-making when compared to the purely analytical methods. Nonetheless, they also point out the fact that the performance of ML is very conditional with regard to the quality and consistency of the data, as environmental data is often characterized by noise, missing points, and uneven distribution of features that may impair the generalization of the models.

The second key ML direction is the application of ML to environmental accounting and Life Cycle Assessment (LCA), in which traditional workflows are constrained by unchanging assumptions, large data demands, and incomplete coverage of lifecycle inventory. Ibn-Mohammed et al. (2023) describe that, trained on lifecycle inventory and sustainability datasets, ML-enhanced LCA can be used to accelerate the assessment, fill gaps, and enhance the estimation of the environmental impact of product and material lifecycles. In the same vein, Adelakun et al. (2024) indicate that the sustainability accounting and monitoring systems that are based on ML are capable of processing heterogeneous indicators, including structured corporate variables and monitoring streams, to enhance transparency, minimize errors in manual reporting, and enable near real-time sustainability reporting. Despite this, both strands of work observe that the reliability of models is limited by the presence of inconsistent data standards, missingness and the difficulty of unifying different data sources into an auditable pipeline.

Classical supervised algorithms have also been used in the estimation of carbon emissions, renewable energy analytics, and industry-specific sustainability forecasting with ML. Sumon et al. (2024) demonstrate that predictive models based on the Random Forest and XGBoost can be trained to perform sustainability-related prediction tasks with strong predictive accuracy when structured datasets are present, although the outcomes are frequently sensitive to careful feature engineering and preprocessing choices. Giannelos et al. (2023) also show that regression-based ML and ARIMA-type methods are useful in providing a baseline of environmental impact predictions over time-series predictors, but their effectiveness may be compromised by highly nonlinear relationships or when there are many predictors interacting with each other. In total, the ML literature justifies the usefulness of algorithmic prediction in sustainability, although the same set of limitations is associated with the heterogeneity treatment, resilience to missing/noisy indicators, and the ability to interpret the influence of features.

2.2 Deep Learning Approaches in Environmental Sustainability

Deep learning (DL) has enhanced the sustainability studies by facilitating the representation learning based on the complex environmental data, especially when the data are multidimensional, spatial, or time-based. As Magazzino (2024) emphasizes, DL systems are more effective in analysis of pollution, remote sensing, biodiversity monitoring, and climate modelling since neural architectures are able to automatically identify meaningful patterns in the sources of information, e.g. satellite images, geospatial data, and sensor streams. Notwithstanding these benefits, compared to classical models, DL systems may be more difficult to interpret, which can decrease trust and usability when predictions are needed to support compliance, policy justification or sustainability auditing.

One of the significant contributions of DL is the time-series forecasting and anomaly detection, where models such as LSTM and GRU are able to learn temporal dependencies that are hard to capture using more traditional regression-based models. Nevertheless, the authors of the article by Martiny et al. (2024) also highlight a concern that is more specific to sustainability: the design of a DL model (layers, neurons, training epochs) not only affects the predictive performance but also the computational energy consumption, which may result in a contradiction when sustainability-related projects are based on the training of energy-intensive models. This underscores the importance of architectures that offer high accuracy and also enhance computational efficiency when models need to be retrained on a frequent basis as the new sustainability data becomes available.

In the industrial and corporate monitoring applications, DL is frequently encompassed with multimodal pipelines which unite sensor, image, and textual data to enhance sustainability of operations and detection of hazards. Wang et al. (2025) demonstrate that multimodal DL systems can be more accurate and responsive than the traditional baselines because they combine various types of data in a single model, which allows modelling more environmental conditions. Likewise, Mu et al. (2023) show that hybrid CNN-LSTM models are able to simulate multi-time-scale environmental cues to enhance monitoring and prediction of corporate ecological performance, but also indicate that the clarity of features at the feature level is still low and the models can be computationally expensive. Together, DL research provides predictive advantages that are robust though still limited in terms of costs, interpretability, and robustness when considering heterogeneous indicators that drive more efficient and feature-conscious models on tabular sustainability data.

2.3 AI-Enabled Environmental Decision Systems for Policy, Finance, and Industrial Innovation

The application of AI within the domain of prediction is not the only aspect of the technology that is being implemented in the decision-making process that would facilitate environmental governance, compliance, and sustainability strategy. Ibn-Mohammed et al. (2023) write about the way in which ML-based LCA methods can enhance the process of policy evaluation through the provision of more data-driven estimates of environmental effects than the purely retrospective ones. Besides, Zhang and Yin (2023) investigate the performance of green innovation concerning policy instruments and demonstrate how predictive and analysis models may be used to measure the impact of administrative penalties, environmental taxes, and pollution-control investments on corporate sustainability achievements. Nevertheless, the use of policy-oriented modelling may be constrained by proxy variables, location variations, and difficulties in extrapolating across sectors in which environmental indicators vary significantly.

In climate finance, AI is taking centre stage in risk models and sustainable investment strategies where the stakeholders need forecasting models that bridges environmental indicators and asset resilience as well as transition risk. Ferrara et al. (2024) maintain that AI-assisted predictive analytics can be enhanced with technologies like blockchain to enhance supply chain and financial network transparency and timeliness of climate risks assessment, which is beneficial in monitoring and accountability. However, these systems have practical constraints on data availability, standardization and complexity of integration especially where institutions adopt disparate reporting format or sustainability measurement standards.

On the industrial and infrastructure level, AI can help with environmental accounting, hazard detection, and cost translation, involving predictive analytics in the working processes. Adelakun et al. (2024) clarify that AI-based sustainability reporting tools can be used to extract, process, and report sustainability indicators, minimizing human error, and providing continuous reporting of corporate reporting. Gao et al. (2025) also demonstrate that AIpowered Building Information Modelling systems are capable of transforming environmental effects into financial cost terms and visualizing them throughout the project phases to facilitate more practical decisions. Nevertheless, interoperability constraints, project or organizational generalization and clear explanations of indicators that yield high predicted environmental costs are common limitations of industrial decision systems.

2.4 Research Gaps

Despite quantifiable advances in sustainability prediction in the application of ML and DL, a number of limitations are recurring in the literature. First, most of the studies rely on datasets which are incomplete, noisy or heterogeneous where missing values and inconsistent measurement practices undermine the predictive reliability and predispose the result to biased outputs (Farooq and Khan, 2025; Ibn-Mohammed et al., 2023). The indicators of sustainability in the real world of corporations and policies sometimes have many different sources and time scales and therefore models that assume clean data that is uniform might not be sustained when applied to non-experimental datasets.

Second, the deep learning methods that enhance predictive capabilities may add high computational costs that are incompatible with the sustainability objectives of training that requires a large amount of energy (Martiny et al., 2024). Moreover, most of the high-performing models do not have good interpretability, and it is hard to justify predictions or which indicators had a significant impact on environmental cost outcomes, which is especially critical to auditability, regulatory compliance, and transparent decision-making (Mu et al., 2023; Zhang and Yin, 2023). The complexity of multimodal approaches additionally adds to the complexity of integration and a significant number of studies lack sufficient implementation descriptions to guarantee that they can be replicated in different organizational settings.

Third, the literature demonstrates the absence of a single standardized benchmarking, in which classical ML, recurrent DL, and feature-aware attention-based tabular models are treated equally under a comparable preprocessing and evaluation process. Such a gap drives methods that are both precise and interpretable of structured and heterogeneous sustainability indicators. In that regard, Sparse-Attention Tab Net is quite suitable as it is tabular data-based and applies sequential attentive masks to achieve adaptive feature selection that can enhance robustness and better show which features have an impact and which do not. Hence, the proposed framework meets the concurrent requirement of the performance, efficiency, and feature-level visibility of environmental costs prediction.

Reference (Year)	**Model / Method**	**Advantages**	**Disadvantage/ Limitations**
Sumon et al. (2024)	Classical ML comparison (RF, XGBoost, regression variants)	ML can model nonlinear sustainability indicators and gives	Performance often depends on heavy preprocessing/feature engineering; lacks attention-

		strong baselines	based feature selection
Zhang & Yin (2023)	Gradient Boosting for corporate sustainability-related prediction	Boosting improves accuracy over simple baselines; works well on structured firm indicators	Still non-attentive and limited in learning feature interactions end-to-end compared with attention-based tabular DL
Mu et al. (2023)	Data fusion + CNN–LSTM for corporate environmental monitoring	Deep learning captures complex nonlinear dependencies in environmental performance signals	Higher computational cost and complex architectures; not tailored for heterogeneous tabular-only indicators
Nazir et al. (2024)	GRU/LSTM/ANN + ensemble forecasting	Recurrent DL improves predictive accuracy compared to basic models	Time-series focus and potential overfitting; not designed for high-dimensional tabular feature selection
Martiny et al. (2024)	LSTM forecasting with training-emissions analysis	Explicitly demonstrates accuracy–compute/energy trade-offs in DL training	Shows sustainability contradiction of compute-heavy training → motivates computationally efficient sparse mechanisms
sAdelakun et al. (2024)	AI for sustainability accounting (ML + NLP, conceptual)	Motivates automated analysis of heterogeneous sustainability indicators	Lacks a benchmarked predictive model; highlights need for scalable, efficient structured-data prediction pipelines
Ibn-Mohammed et al. (2023)	AI/ML-enabled LCA enhancement framework	Supports shift from static LCA to predictive, data-driven estimation	Emphasizes data gaps and heterogeneity; limited empirical benchmarking → motivates robust tabular model with adaptive feature learning

3 Research Methodology

3.1 Methodology

This paper applies CRISP-DM to create a Sparse-Attention TabNet to address the following common drawbacks of environmental analytics: Working with heterogeneous multi-source data, missing values, and the high cost of dense deep learning. During Business Understanding, the aim is to create an effective and consistent predictive model on a variety of environmental data, whereas in Data Understanding, datasets that include climate indicators, renewable energy metrics, policy variables, and corporate environmental performance metrics are analysed in terms of distribution, scale variations, missingness, and complexity to verify suitability to TabNet. The Data Preparation and Modelling stages are inherently compatible with TabNet, cleaning data, encoding features, handling missing values, and formatting inputs to take attention-based steps in decision making, and setting sparse attention to trim down unnecessary computation, noise reduction, and resilience to different environmental conditions. Lastly, under Evaluation and Deployment, the model is evaluated on the basis of accuracy, efficiency and stability to make sure that it can be reliably used in environmental monitoring, climate assessment, and sustainability-oriented prediction tasks with CRISP-DM providing a systematic, practical end-to-end development cycle.

3.2 Dataset Description

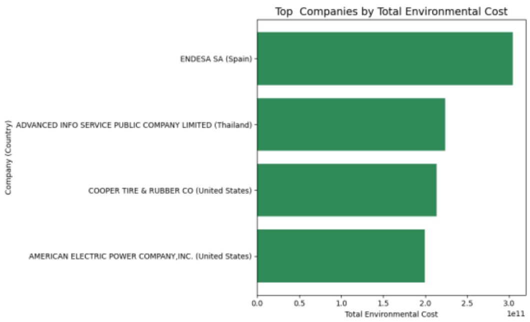
The data sample comprises firm-level indicator of environmental performance of businesses operating in the sector of Activities auxiliary to financial intermediation in several countries in 2018. It has quantitative indicators of sustainability modelling including environmental intensity in terms of revenue and operating income as well as overall environmental cost. There are also operational and sector-linked capacity variables, which consist of working capacity and production capacities of fish, crops, meat, wood, and water. Moreover, the dataset includes ecological pressure in terms of such indicators as the loss of biodiversity and depletion of abiotic resources. It also incorporates imputed percentages estimates which are either uncertainty or estimated elements in the data. In general, the data is mixed with heterogeneity, high-dimensional, negative-valued, high magnitudes of costs, and different levels of missingness. These features are indicative of general issues of environmental data and previous sustainability research. Due to this complexity, the data is well-aligned with Sparse-Attention TabNet which has the ability to process mixed scale metrics, minimize noise, and learn important relationships among a wide range of indicators.

3.3 Data Preprocessing

The preprocessing began with the importation of the dataset and its observation by examining the data representation, the number of rows and columns, and the types of data. This exposed more than 13,000 records containing a combination of text, categorical and numerical fields with a large number of numeric values being represented as strings due to the presence of commas, bracketed negatives and percentages. It was verified that the data is complete and not missing and cleaned by ensuring that it does not contain commas, negative formatted values are converted to valid numeric values, and percentage values are converted to decimals as this will leave all the relevant columns as valid continuous variables that can be used to train a model.

Then, blank entries in a few variables like the intensity of operating-income, the abiotic resources, and wood production were replaced with the median of each column to maintain the overall distribution and eliminate the distortion of the extreme values. Categorical variables such as company identity, country, and industry classification had their integer encoding to numeric labels so that the learning model can effectively process them like the continuous environmental indicators. The data was cleaned and coded and divided into training and testing subsets with an eighty-twenty proportion to facilitate learning and fair assessment. Then, the continuous input features were standardized to have the same scale, which enhanced their training stability and convergence. Environmental cost is also considered as the target variable and it was also standardized in the same way and the same transformations were performed in both the training and testing sets.

3.4 EDA (Exploratory Data Analysis)



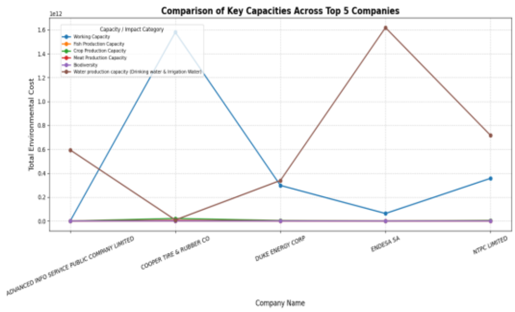
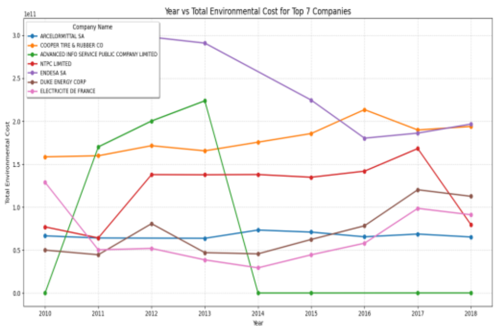


Figure 1. Companies by Environmental Cost **Figure 2. Capacity Comparison of Top Firms**

Plot in figure 1 emphasizes the companies with the highest total environmental cost in the process of which the ENDESA SA (Spain) has the highest overall impact than other top contributors. The plot in figure 2 compares various environmental capacities of the top five companies and shows that there is a lot of difference in the working capacity, water-related impacts, and biodiversity pressures among the firms. Collectively, these visualizations demonstrate that environmental cost is not just high in relation to specific companies but also it is determined by various capacity and resource-related issues.



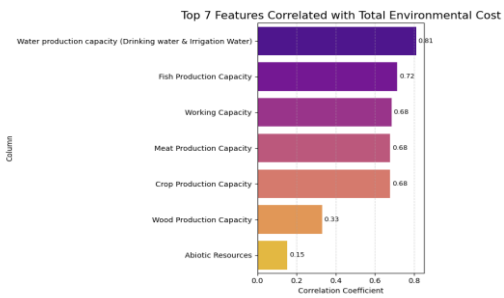


Figure 3. Environmental Cost Over Time **Figure 4. Top Cost Correlations**

Figure 3 illustrates total environmental cost versus time of the top seven companies, and the companies such as ENDESA SA and COOPER TIRE and Rubber Co are high and varying with varying years. Figure 4 indicates the most positive connections to environmental cost with water and fish production capacity at the forefront of other measures of operational capacity.

3.5 Sparse-Attention TabNet Model and Advantages

To overcome the limitations of the available literature, the proposed study will seek to suggest and utilize a Sparse-Attention TabNet model to mitigate the data heterogeneity, computational inefficiency, and the inability to understand the features inherent in the existing environmental sustainability literature. Older ML models did not support the ability to identify meaningful relationships between mixed environmental data, whereas deep learning models were associated with high computational cost, lack of visibility in the internal processes, and the challenge of dealing with missing or imbalanced data. Sparse-Attention TabNet model is present to address these problems with the help of sequential attention-based decision-making steps during which the most significant environmental features are selectively focused on. Its internal sparsity algorithm minimizes redundant calculations and only significant traits of predictions are emphasized, which makes it consume more and less energy. This leverage model is also superior in terms of dealing with complex tabular environmental data with minimal preprocessing thus it can in effect combine climate variables, policy indicators, renewable energy metrics and corporate environmental performance measures within a single analytical framework.

The benefits of fusing sparse attention with TabNet are also seen in the increased model efficiency, increased robustness, and increased relevance to real-world environmental tasks. Sparse attention provides this property such that the model only pays attention to the most informative signals among a variety of datasets, reducing noise sensitivity and enhancing generalization to a variety of environmental contexts. The decision-step architecture of TabNet offers intuitive explanations of the predictions, making its use accessible to policymakers, sustainability officers, and researchers to understand which environmental conditions, or a combination of environmental conditions, has the largest effect on climate risk, its performance metrics, or sustainability outcomes. Also, the efficiency of the model when using a smaller computation cost allows its use in large-scale environmental monitoring tasks, which is relevant to the carbon-emission issues related to the deep learning training processes. Combining these features, the Sparse-Attention TabNet framework is a strong, interpretable, and energy-efficient analytical framework that directly addresses the fundamental gaps that were found in previous research.

3.6 Tab Net Model Architecture and Operational Mechanism

The architecture of TabNet in this research applies to the complex data of the environment in a straightforward sequence of steps as shown in figure 5. The raw training data, including climate indicators, emission, renewable energy statistics, and corporate sustainability variables, are initially processed by the Feature Processor that converts them into a consistent representation and aligns them without any manual feature selection. This step will be very appropriate in environmental datasets, as they tend to have mixed formats, non-uniform scales and incomplete or noisy records, therefore the model will advantage in a structured manner of preparing all signals to be learned.

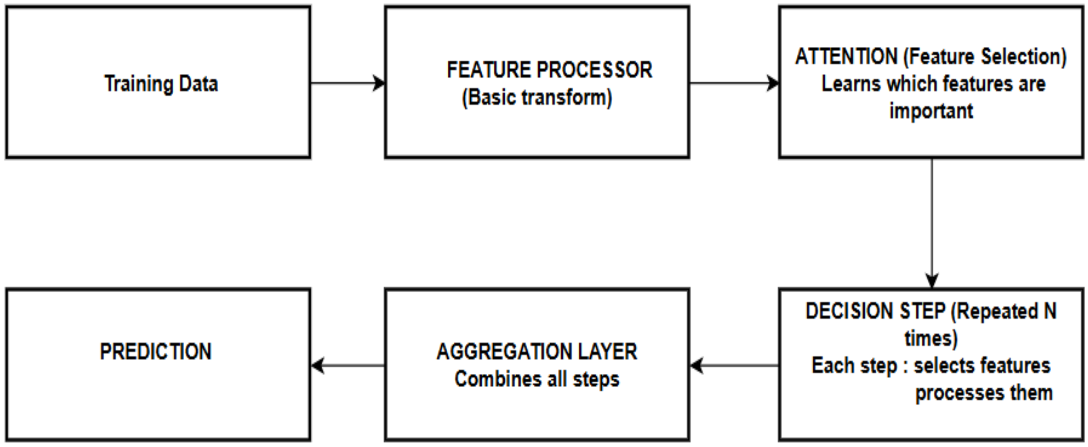


Figure 5. Overall Workflow of the TabNet Model Architecture

The transformed features then enter the Sparse-Attention layer which generates attention masks to put emphasis on the most significant environmental variables in each learning step to eliminate redundancy and noise in correlated indicators. These chosen features undergo a series of Decision Steps that narrow the focus of features repeatedly, and draw more out of them, as the model learns how joint influences of factors like energy adoption, industrial activity and regulation determine sustainability results. Lastly, the Aggregation layer combines all the decisions steps into one representation, and the Prediction module generates such outputs as environmental cost or sustainability performance, which provides meaningful and understandable results.

3.7 Model Evaluation Metrics

Four popular regression measures, such as Mean Squared Error, Root Mean Squared Error, and R2 Score are used to evaluate the performance of TabNet model in predicting environmental cost, although MSE focuses more on the large errors and thus it is useful in determining whether the model is sensitive to the extreme values. RMSE is the square root of MSE, which presents error in the same units as the target variable which are easy to interpret. R2 Score: It is used to show the amount of variance in environmental cost that the model explains relative to a baseline. Combined, the metrics provide a clear evaluation of the accuracy of prediction, the severity of errors, and explanatory power on sustainability modelling.

4 Design Specification

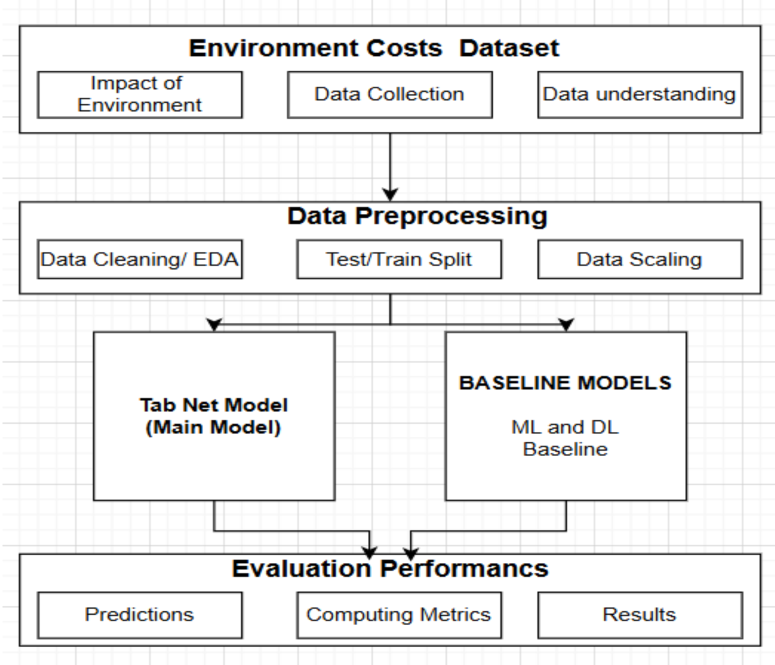


Figure 6. End-to-End Pipeline for Environmental Cost Prediction

Figure 6 is a prediction framework of the environmental cost that integrates Machine Learning, Deep Learning, and TabNet into a single pipeline, which transforms raw environmental data into verified sustainability predictions. It is based on Dataset, Preprocessing, and Modelling/Evaluation modules, TabNet is the primary model of sparse-attention tabular learning and ML/DL models.

4.1 System Architecture Overview

The environmental cost prediction system architecture is designed into four layers that are interconnected, and they include: Environmental Dataset Layer, Data Preprocessing Layer, Model Development Layer, and Evaluation Layer. The dataset layer helps to obtain and gain the initial insights into the environmental indicators at the firm level, such as production capacities, environmental pressures, and environmental metrics costs. This layer provides the modelling process with the contextual basis or foundation because it makes certain that the relevant environmental inputs are captured and interpreted systematically. The next data preprocessing level normalizes and prepares the data with cleaning, converting to numbers, addressing missing data, encoding discrete data, and normalizing continuous data. This will make sure that the data is formatted in a consistent manner, analytically consistent and prepared to be consumed by various machine learning and deep learning models.

The Model Development Layer is a framework that combines traditional machine learning models, deep learning architectures, and the TabNet model into one analytical model. Both models are fed the processed data but utilize varying learning processes such as linear regression and ensemble learning to neural feature extraction and sparse-attention transformations. The TabNet architecture is the primary model because it is specifically designed to work with tabular data and it can prioritize the important environmental attributes by sequential attention. Lastly, the Evaluation Layer calculates the predictive measures, compares the performance of the models in different categories and delivers its results that will be used in the sustainability analysis. The architectural layers together are scalable, modular and transparent workflow to tackle the complexities of heterogeneous environmental datasets.

4.2 Pipeline Design for Model Training

The model-training pipeline starts with an organized analysis of the environmental data, in which the system validates the format of the variables, the presence of missing data, the type of data, distribution patterns of the variables to detect any anomalies like the numbers represented as strings or values with special characters. These diagnostics inform the cleaning step that operates under numeric conversion, deletion of undesired characters, median imputation of missing values, categorical variable label encoding, and z-score normalization of continuous variables and then splits the data into 80/20 train/test. With preprocessing, several models are trained such as ML models, such as Decision Tree, Random Forest, Lasso, and SVM, which capture nonlinear relations, and DL models, such as MLP, DNN, LSTM, and GRU, which optimize gradients, use ReLU activations, and dropout, and recurrent models which are fed on reshaped pseudo-temporal inputs in order to learn structural dependencies. Lastly, TabNet architecture handles tokenized and normalized features with its sparse-attention model where attentive transformers will select the relevant variables, and feature transformers will learn nonlinear patterns among sequential decision-making steps, which provides the model with the ability to capture all the complex hierarchical interactions among environmental indicators and generate correct sustainability predictions.

4.3 Tab Net Architecture

Implementing Tab Net model was based on TabNet Regressor framework which uses a sequential attention mechanism with tabular environmental data allowing the model to dynamically choose the most informative features at each decision step. The architecture was set up to use a 16-dimensional decision and attention embedding, five decision steps, a sparsity regularization coefficient, and the sparsemax mask function to implement and focused feature selection. The input attributes were transformed to a projection-based feature transformer and then subjected to a series of attentive steps, with each step being trained to learn a different set of environmental indicators depending on the sparse attention masks of the model. The 200 epochs were trained on Adam optimizer with learning rate of 0.02, with large batch and virtual batch size to aid in stabilization of the optimization. Monitoring performance in terms of RMSE metrics was done in a separate validation set during training. The model after convergence produced predictions that were the sum of output of all decision steps and reflected non-linear interactions among climate variables, production capacities, ecological pressures and environmental cost indicators. This application enabled TabNet to capitalize on its natural computational efficiency, and selective feature-learning features, and it is suitable to complex sustainability modelling problems.

4.4 Tab Transformer Model Implementation

The Tab Transformer model was applied through an architecture of transformer based on structured environmental data. A linear projection of the input features with ReLU activation and layer normalization to an embedding of 64 dimensions was utilized as the first step in the process. These embeddings then passed through two Stable Transformer Blocks which included multi-head self-attention with four heads, residual connections and dropout to help learn cross-feature interactions and also to keep the training stable. The output representations were then pooled after processing the attention and then fed through a multilayer perceptron consisting of 128 hidden units and dropout, then a linear layer was used to produce the final environmental cost prediction. The model was trained over 200 epochs with Adam optimizer with a learning rate of 0.0003 and batch size of 16 so that it could learn the dependency of features in the transformed tabular space as shown in figure 7.

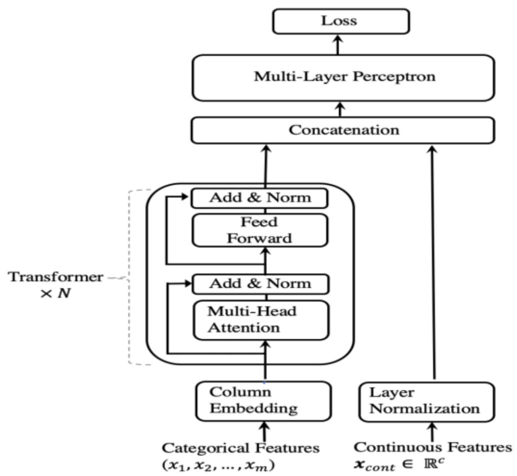


Figure 7.Internal Architecture of Tab Transformer

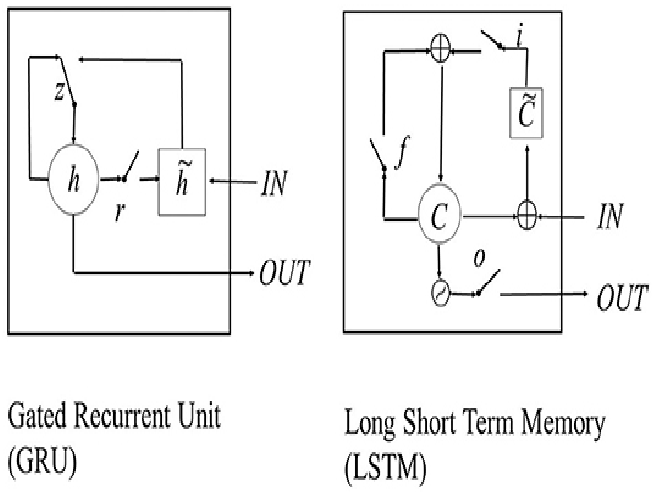


Figure 8. Internal Architecture of LSTM and GRU

4.5 Deep Learning Models Implementation

The Multi-layer Perceptron (MLP) model was developed based on a single-hidden-layer neural network with 16 neurons and ReLU activation, which was trained with the Adam solver and adaptive learning rate scheme. The training was done in an iterative fashion and with 16 epochs and warm-start to enable the model to keep on optimizing its weights throughout the successive fitting cycles. This structure allowed the MLP to acquire non-linear correlations between the scaled environmental features and also to remain stable during gradient updates. The model was able to gradually minimize reconstruction error and learn patterns associated with environmental cost prediction due to the network structure, coupled with the epoch-based training.

The Deep Neural Network (DNN) was built as a multi-layer network comprising of dense layers with 64, 32 and 16 neurons respectively and each layer incorporates ReLU activation to enhance extraction of non-linear features. The first and second hidden layers were followed by dropout layers to avoid overfitting by randomly deactivating neurons during training. The model was trained in 40 epochs and a batch size of 256 with Adam optimizer and a learning rate of 0.0001, which guaranteed a smooth gradient flow and a stable convergence. The additional architecture enabled the DNN to learn hierarchical feature representations on the environmental dataset and consequently, capture complex variable interactions more effectively in contrast to shallower networks.

The neural network based on GRUs was aimed at exploiting gated recurrent units to process structured sequences based on tabular features reconfigured to a pseudo-temporal representation. The model had a GRU layer of 64 units and tanh activation, then 32 and 16-neuron dense layers that were backed by dropout regularization. The training was done by Adam optimizer at a learning rate of 0.0001 and 40 epochs with a batch size of 256. Even though the data was not in itself sequential, the GRU architecture allowed the model to learn inherent associations between environmental cues using its gating mechanism and therefore, it was capable of learning interactions between features in a similar way to temporal sequence modelling.

The LSTM based model shared the same structural design with the GRU architecture, i.e. having an LSTM layer with 64 units and tanh activation that would be used to capture the long-range dependency between the reshaped environmental features. There were dropout layers with increased dropout rates (0.4 and 0.3) to minimize the overfitting risks since the LSTM has more parameters. Later layers of dense coupled with ReLU activation enabled pattern extraction through hierarchy and the last linear output layer made environmental cost predictions. The model was trained on Adam optimizer with a learning rate of 0.0001 and 40 epochs and 64 by batch size, which allowed learning the feature-level dependence that was essential to the sustainability metrics as shown in figure 8.

5 Implementation

5.1 Machine Learning Models Implementation

The machine learning models in this study were optimized by hyperparameter tuning, which is a cross-validated search method, in which parameter ranges were predefined and the most effective set of parameters was evaluated on each model. The tuning was done on parameters that affect model complexity, regularization strength and behaviour of the kernel, so that every best estimator would be chosen based on its capacity to generalize to unseen environmental data.

In the case of the Decision Tree Regressor, the tuning process tested the parameter minsamples split with the following values: is 2, 5, 10, 20 and this is the minimum number of samples required to split an internal node. Smaller values will enable the tree to make more detailed partitions whereas larger values will limit unnecessary branching. Among the tested alternatives, the most suitable model of minsamplessplit has the best performance of 2, so that the more splits were allowed, the more the dataset predictive structure. This best estimator was effective in fitting nonlinear trends in the environmental cost variables without adding too much tree depth as in figure 9.

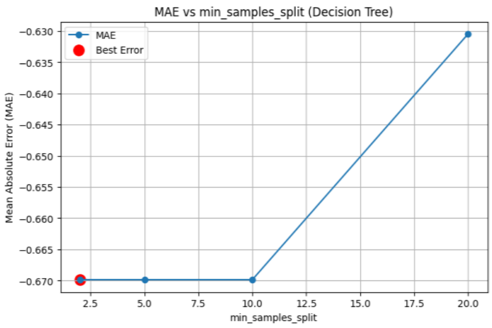


Figure 9. DT Error for Split Settings

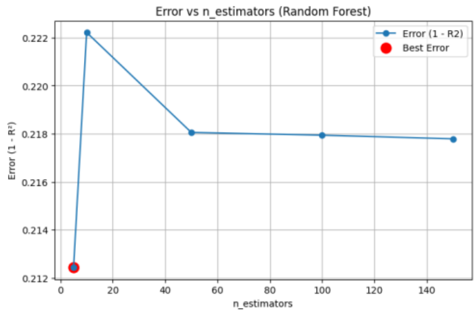


Figure 10. Error with No of estimators RF

Estimators were the main parameter that was investigated in the Random Forest Regressor and the values of is 5, 10, 50, 100, 150 to determine the impact of ensemble size on the model performance as illustrated in figure 10. A lower count of trees will tend to be quicker in calculation, but larger ensembles will have a more consistent and even prediction. The best estimator was achieved using number of estimators among the tested configurations as 5 implying that a small ensemble could be used to model the underlying relationships in the environmental data without compromising on the computational efficiency.

In the case of the Lasso Regression model, the tuning was using the regularization parameter alpha that was tuned with values of 0.3, 0.5, 0.7, 0.9. This parameter directly determines the strength of shrinkage of the coefficients, which encourages sparsity by diminishing the impact of insignificant features. Among these alternatives, the most optimal estimator was associated with alpha = 0.3, the lowest value tested, which suggested that a small amount of regularization was the most suitable to balance between having important environmental predictors and overfitting in the linear model.

The parameters to be tested in the Support Vector Regression (SVR) model were C 1, 10, 100, 150, 200 and gamma. C is used to regulate the penalty assigned to errors and gamma is used to regulate the curvature on the decision boundary when the RBF kernel is used. The most successful estimator, which was C200 and gamma, showed that a strong penalty with smooth kernel was the most effective model in the process of capturing nonlinear variation in the environmental cost data.

6 Evaluation

Table 1. Performance Comparison of Machine Learning Models

Model	R2 Score	MSE	RMSE
Decision Tree	0.82	2432.82	49.32
Random Forest	0.88	1623.38	40.29
Lasso	0.84	2078.72	45.59
SVR	0.79	2793.61	52.85

Table 2.Performance Comparison of Deep Learning Model

Model	R2 Score	MSE	RMSE
MLP	0.91	1117.20	33.42
DNN	0.92	999.12	31.60
GRU	0.95	638.16	25.26
LSTM	0.95	609.60	24.69
Tab Transformer	0.96	435.30	20.86
Tab Net Model	0.98	149.59	12.23

In this part, several ML, DL and transformer-based models are evaluated on the basis of environmental cost prediction in terms of R2, MSE and RMSE as is presented in Tables 1 and 2.

6.1 Machine Learning Model Results

The Decision Tree Regressor gave a R2 score of 0.82, Mean Squared Error of 2432.82 and a Root Mean Squared Error of 49.32. These findings indicate that the model was capable of explaining a significant amount of environmental cost data variance, but the relatively large and RMSE values suggest that the tree structure could have had a difficulty with nonlinear trends, which resulted in larger prediction errors. Although it's the Decision Tree had a low generalization potential, characteristic of shallow or constrained tree structures, it was low. The Random Forest Regressor has got R2 of 0.88, MSE of 1623.38 and RMSE of 40.29. These findings support the claim that the ensemble strategy was effective in overfitting and modelling more subtle relationships in the data. The reduced values of error and high explanatory power are indicative of the strength of the model and its ability to model the complex environmental cost interactions among a wide range of indicators.

The Lasso Regression model gave an R2 of 0.84, MSE of 2078.72 and RMSE of 45.59. Despite the fact that Lasso outperformed the Decision Tree, the fact that Lasso has a larger MSE and RMSE than Random Forest implies that it is incapable of modelling nonlinear relationships because the algorithm is linear. Nonetheless, it is relatively low, which means that Lasso was able to capture important linear relationships, as it took advantage of coefficient shrinkage to eliminate noise and overfitting in the high-dimensional environmental data. The Support Vector Regression model resulted in an R2 of 0.79 having MSE of 2793.61 and RMSE of 52.85. Increased RMSE and decreased R2 characterizes the inability of SVR to handle larger deviations and describe a significant amount of total variance. This discrepancy indicates that SVR is sensitive to scaling, kernel parameters, and intricate changes in environmental dataset, which leads to the variation in the performance measures.

6.2 Deep Learning Model Results

The model MLP had an R 2 = 0.91, Mean Squared error = 1117.20 and RMSE = 33.42. These values show that MLP was able to learn nonlinear patterns more efficiently than the traditional machine learning models as it produced significantly smaller errors. The large R2 indicates that the network could learn more complicated dependencies among environmental indicators, but the error level was a bit larger compared to the deeper architectures. Deep Neural Network (DNN) also enhanced the performance, showing an R2 value of 0.92, a MSE of 999.12 and RMSE of 31.60. The general reduced MSE and RMSE show a high degree of stability and decreased variation in the prediction. The multiple dense-layer architecture facilitated the model to extract hierarchical interactions of features, which allowed the model to explain a higher percentage of variance in environmental cost data.

GRU model showed a considerable performance improvement since the R2 score was 0.95 with an MSE of 638.16 and an RMSE of 25.26. These findings indicate that the GRU was useful in using its gating mechanism to capture dependencies between reshaped environmental features. This significant decrease in RMSE over fully connected networks indicates that the GRU learned more stable representations and extrapolated more across the data set. The LSTM model had the best performance of the deep learning architecture with R2 of 0.95, MSE of 609.60 and RMSE of 24.69. The prediction accuracy and the maximum strength of explanation were achieved because of the model being able to capture long-range feature relationships using memory cells. Its greater performance demonstrates the competence of the sequence-based modelling methods, even in cases of the application to pseudo-temporal transformations of tabular environmental information.

6.3 Tab Transformer Model Results

Tab Transformer model provided good predictive results, and the R2 of the model was 0.96 with a MSE of 435.30 and an RMSE of 20.86. These findings demonstrate that the transformer-based architecture is able to effectively learn the complex interactions of features in the environmental dataset and perform better than traditional and deep learning models in accuracy and variance explanation. The values of the low error and high R2 denote that the Tab Transformer has been able to utilize the attention mechanisms to obtain richer representations of tabular sustainability data.

6.4 TabNet Model Results

The Tab Net model performed the best since it has an R2 score of 0.98 with an MSE of 149.59 and RMSE of 12.23, and it is the most suitable machine learning, deep learning, and transformer-based model when compared to the rest of the models tested in this paper. These very positive measures indicate that Tab Net can learn rich and discriminatory representations on complex environmental data with the help of its sequential attention mechanism. In contrast to traditional neural networks, Tab Net is also selective in identifying the most important features in each step of a decision, enabling it to approximate complex nonlinear relationships and removing noise in less informative variables. The fact that the prediction error has been significantly lowered, and the variance explained is nearly perfect, shows that Tab Net has been able to capture the underlying structure of the environmental cost patterns. Besides, its sparsity-based architecture allows facilitating more explicit definition of the environmental indicators which had the strongest impact on the model predictions. In general, Tab Net model was characterized by better accuracy, robustness, and feature-level transparency, which highlights its appropriateness in sustainability analytics and high-dimensional environmental modelling.

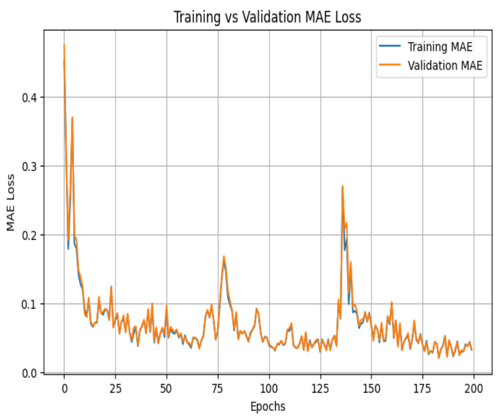


Figure 11. Training vs Validation Loss of Tab Net Model

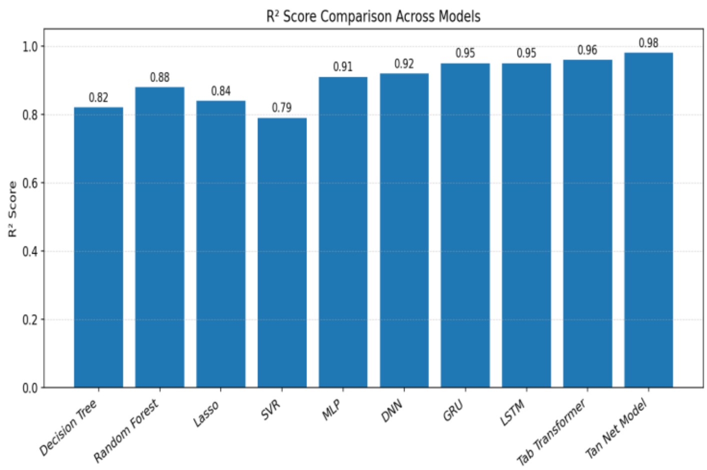


Figure 12. Comparison Models with R2 score

The following plot (figure 11) compares the Training and Validation with 200 epochs where the both the losses decline sharp in the beginning, and level off slowly as learning progresses. The two curves are close together which means that there is consistency in the learning process with the model training.

6.5 Findings of the Study and Model Superiority

In figure 12, the results of this research are clear evidence of the high-performing TabNet model in comparison to all the machine learning, deep learning, and transformer-based architectures tested. The best predictive accuracy and the lowest values of errors obtained by TabNet in all measures were largely due to the unique sequential attention-based learning mechanism that enables the model to selectively concentrate on the most informative environmental features at every decision step. In contrast to conventional neural networks, which do not analyse the features at the same time, TabNet uses sparse feature masks, which reduce noise, avoid overfitting, and reveal the variables most valuable to the prediction of environmental costs. This selective attention together with the capability to learn nonlinear interactions allows TabNet to learn intricate dependencies between climate indicators, production capacities, ecological pressures, and corporate environmental features better than MLPs, DNNs, LSTMs, or GRUs. Also, the architecture of Tab Net makes the learning process transparent, making it possible to obtain meaningful information about the importance of features and at the same time being computationally efficient. All these architectural strengths allow explaining why the TabNet model achieved extraordinarily high accuracy and near-perfect variance explication and can be considered the most powerful and reliable predictive framework in this sustainability-oriented analysis.

7 Conclusion and Future Work

In this research, a full environmental cost prediction framework based on machine learning, deep learning, transformer-based models, and the latest TabNet architecture are used to examine diverse sustainability datasets. The study has shown that Tab Net sparse-attention decision mechanism is able to provide better predictive accuracy and explanatory power than any other model investigated, by overcoming major issues in environmental analytics like heterogeneous feature structures, missing values, nonlinear dependencies. The systematic CRISP-DM framework, the strict preprocessing pipeline, and multi-model comparative appraisal were a strong base to evaluate the performance of the algorithms on various environmental indicators. The results indicate that TabNet has a distinct advantage of determining influential features, noise reduction, and modelling complex environmental relationships at low computational cost, which makes it a successful next-generation instrument in corporate sustainability evaluation, climate-related decision-making, and environmental performance modelling. Altogether, this research paper adds a scalable and very accurate analytical methodology that can enhance data-driven sustainability research and help to make more informed environmental planning.

This study can be further expanded in future studies by examining bigger and more varied environmental data sets and include more socioeconomic, climate, and policy related variables to increase the reliability of predictions. The incorporation of multimodal data like satellite visuals and geospatial layers and real-time sensor feeds can also enhance the capability of the model to integrate dynamic environmental trends. The enhancement can be done by trying the hybrid attention-based architecture or optimization of TabNet to achieve low computational cost and high accuracy. Also, the use of explainability models like SHAP or Integrated Gradients may provide more information on the importance of features and assist sustainability professionals in decision-making. Lastly, the implementation of the model in actual monitoring systems or integration with corporate environmental dashboard would offer a chance to test its practical applicability and long-term functionality in the work environment.

References

Adelakun, B.O., Antwi, B.O., Ntiakoh, A. and Eziefule, A.O., 2024. Leveraging AI for sustainable accounting: Developing models for environmental impact assessment and reporting. Finance & Accounting Research Journal, 6(6), pp.1017-1048.

Bianchi, O. and Putro, H.P., 2024. Artificial intelligence in environmental monitoring: Predicting and managing climate change impacts. International Transactions on Artificial Intelligence, 3(1), pp.85-96.

Crippa M, Guizzardi D, Banja M, Solazzo E, Muntean M, Schaaf E, Pagani F, Monforti-Ferrario F, Olivier J, Quadrelli R, Risquez Martin A. CO2 emissions of all world countries. JRC Science for Policy Report, European Commission, EUR. 2022;31182.

Farooq, O. and Khan, A., 2025. The Impact of Machine Learning on Climate Change Modeling and Environmental Sustainability. Artificial Intelligence and Machine Learning Review, 6(1), pp.8-16.

Ferrara, M., Ciano, T., Capriotti, A. and Muzzioli, S., 2024. Machine Learning Predictive Modeling for assessing Climate Risk in Finance. WSEAS Transactions on Environment and Development.

Gao, X., Chen, X., Lu, K. and Deng, X., 2025. Monetizing Environmental Impacts into Environmental Costs During Prefabricated Building Construction: A 5D BIM-Enabled Analysis. CivilEng, 6(3), p.36.

Gazi, M.S., Barua, A., Karim, F., Siddiqui, M.I.H., Das, N., Islam, M.R. and Al Montaser, M.A., 2025. Machine Learning-Driven Analysis of Low-Carbon Technology Trade and Its Economic Impact in the USA. Journal of Ecohumanism, 4(1), pp.4961-4984.

Giannelos, S., Moreira, A., Papadaskalopoulos, D., Borozan, S., Pudjianto, D., Konstantelos, I., Sun, M. and Strbac, G., 2023. A machine learning approach for generating and evaluating forecasts on the environmental impact of the buildings sector. Energies, 16(6), p.2915.

Ibn-Mohammed, T., Mustapha, K.B., Abdulkareem, M., Fuensanta, A.U., Pecunia, V. and Dancer, C.E.J., 2023. Toward artificial intelligence and machine learning-enabled frameworks for improved predictions of lifecycle environmental impacts of functional materials and devices. MRS Communications, 13(5), pp.795-811.

Magazzino, C., 2024. The impact of deep learning on environmental science. BMC Environmental Science, 1(1), p.4.

Martiny, A., Meo, M. and Vallero, G., 2024, October. Analysis of LSTM Networks for Reduced Environmental Impact in Time Series Forecast. In 2024 20th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (pp.148-153). IEEE.

Mu, Y., Duan, C., Li, X. and Wu, Y., 2023. A Monitoring Method for Corporate Environmental Performance Based on Data Fusion in China under the Double Carbon Target. Sustainability, 15(12), p.9391.

Nazir, T., Nagra, H.F., Bhatti, M.H., Shahid, R.N. and Shaukat, N., 2024. Predictive analytics for environmental impact: Deep learning and AI ensemble strategies in forecasting CO_2 emissions.

Prashanth, P., Elmourad, J., Grobler, C., Isaacs, S., Zahid, S.S., Abel, J., Falter, C., Fritz, T., Allroggen, F., Sabnis, J.S. and Eastham, S.D., 2024. Near-zero environmental impact aircraft. Sustainable Energy & Fuels, 8(20), pp.4772-4782.

Sumon, M.F.I., Osiujjaman, M., Khan, M.A., Rahman, A., Uddin, M.S., Pant, L.P. and Debnath, P., 2024. Environmental and socio-economic impact assessment of renewable energy using machine learning models. Journal of Economics, Finance, and Accounting Studies, 6(5), p.112.

Wang, S., Cheng, N. and Hu, Y., 2025. Comprehensive environmental monitoring system for industrial and mining enterprises using multimodal deep learning and clip model. IEEE Access.

Zhang, J. and Yin, K., 2023. Application of gradient boosting model to forecast corporate green innovation performance. Frontiers in Environmental Science, 11, p.1252271.