Project Synopsis

on

**Project Synopsis on A Specialized Multi-Model Machine Learning Framework for Granular Carbon Emission Forecasting in the Mining Sector**

Submitted as a part of course curriculum for

Bachelor of Technology

in

Computer Science and Engineering



Submitted by

Sahil Tyagi(2200290100138)

Sahil Kumar(2200290100137)

Rohan Singh(2200290100136)

Rohan Singh(2200290100135)

Under the Supervision of Dr. Deepti Singh

Associate Professor

Department of Computer Science and Engineering­

**KIET Group of Institutions, Ghaziabad Department of Computer Science**

**Dr. A.P.J. Abdul Kalam Technical University 2024-25**

**ACKNOWLEDGEMENT**

It gives us immense pleasure to present the synopsis of our B.Tech Major Project undertaken during the third year of our Bachelor of Technology program. We express our heartfelt gratitude to **Dr. Deepti Singh**, Associate Professor, **Department of Computer Science and Engineering**, KIET Group of Institutions, Delhi-NCR, Ghaziabad, for her invaluable guidance and continuous support throughout the course of this project. Her dedication, insightful feedback, and encouragement have been a driving force in the successful completion of our work.

We are also deeply thankful to **Dr. Vineet Sharma**, Dean, Department of Computer Science and Engineering, KIET Group of Institutions, Delhi-NCR, Ghaziabad, for his constant encouragement and support during the development of our project.

We would also like to extend our sincere thanks to all the **faculty members of the department** for their valuable suggestions, guidance, and support at various stages of the project.

Last but not the least, we acknowledge the contribution and cooperation of our **friends and peers**, who directly or indirectly helped us complete this project successfully.

Signature:

Student’s Name: Sahil Tyagi

Roll No: 2200290100138

Signature:

Student’s Name: Sahil Kumar

Roll No: 2200290100137

Signature:

Student’s Name: Rohan Singh

Roll No: 2200290100136

Signature:

Student’s Name: Rohan Singh

Roll No: 2200290100135

Guide Name & Signature

#### Dr. Deepti Singh

Associate Professor Department of

Computer Science and Engineering

KIET Group of Institutions

# ABSTRACT

The Indian mining industry, a critical component of the national economy, faces a significant challenge in managing and mitigating its carbon footprint. Current methods for carbon accounting often depend on generalized, high-level estimations, which lack the required granularity for effective, real-time operational decisions. This project directly addresses this gap by proposing and developing a novel, web-based platform that utilizes a **specialized multi-model machine learning framework** for the precise forecasting and management of carbon emissions.

Our innovative approach moves beyond a singular, "one-size-fits-all" predictive model. We architect a system of four distinct "expert" machine learning models, each meticulously trained to predict emissions from a specific operational domain: **1) Electricity Consumption, 2) Transportation Logistics, 3) Fuel Combustion, and 4) Explosives Usage.** This modular architecture allows the system to capture the unique emission signatures of each activity with high fidelity.

The technical core of the system is built using **Random Forest Regressors**, an advanced ensemble algorithm chosen for its high accuracy and robustness. The platform will not only provide precise quantitative forecasts in kg of CO2 but will also translate these complex outputs into an intuitive, qualitative **Risk Level** (Low, Medium, High, Very High), enabling at-a-glance operational awareness for mine managers. The primary contribution of this work is a framework that delivers not just prediction, but inherent **source attribution**. By pinpointing the exact operational drivers of emissions, our tool empowers the mining industry to transition from reactive monitoring to proactive, data-driven carbon management.

# TABLE OF CONTENTS

|  |  |
| --- | --- |
|  | Page No |
| TITLE PAGE .................................................................................................................... | i |
| ACKNOWLEDGEMENT................................................................................................. | ii |
| ABSTRACT...................................................................................................................... | iii |
| CHAPTER 1 - INTRODUCTION………………………………………………………. | 1 |
| CHAPTER 2 - LITERATURE REVIEW……………………………………………...... | 2-3 |
| CHAPTER 3 – PROBLEM STATEMENT……………………………………………... | 4 |
| CHAPTER 4 – OBJECTIVES…………………………………………………………... | 5 |
| CHAPTER 5 - PROPOSED SYSTEM …………………………………......................... | 6-8 |
| CHAPTER 6 – MODULES AND IMPLEMENTATION…….………………………... | 9-10 |
| CHAPTER 7 – EXPECTED RESULTS AND PERFORMANCE ANALYSIS ……. | 11 |
| CHAPTER 8 - CONCLUSION & FUTURE SCOPE ………………………………….. | 12 |
| REFERENCES….............................................................................................................. | 13 |

**CHAPTER 1 - INTRODUCTION**

The global imperative to address climate change has placed significant and escalating pressure on heavy industries to curtail their greenhouse gas (GHG) emissions. For India, a nation balancing rapid economic growth with profound environmental responsibilities, the coal mining sector represents a complex challenge. It is both a vital engine of industrial and economic activity and a major contributor to the nation's overall carbon footprint. To align with national commitments, such as the Paris Agreement, and to navigate an increasingly carbon-conscious global market, it is essential for Indian mining companies to adopt sophisticated, 21st-century tools. These tools must evolve beyond simple, retrospective accounting and provide forward-looking, actionable intelligence for tangible emission reduction .

This project is born from the recognition that effective environmental management is fundamentally a data science problem. The operational environment of a coal mine is a complex system of interconnected processes—from the immense energy consumed in excavating and processing materials, to the logistics of a vast transportation network, to the chemical reactions of fuel combustion and explosives. Each of these activities has a unique and dynamic "emission signature." Our project introduces a cutting-edge machine learning framework designed to decode these signatures, providing the high-resolution insights necessary for strategic, data-driven interventions.

# CHAPTER 2 - LITERATURE REVIEW

Our research is built upon a comprehensive review of existing work in carbon emission modeling, machine learning, and mining operations. The following papers form the core of our literature foundation:

**[1] Kumari, S., & Singh, S. K. (2023). Machine learning-based time series models for effective CO2 emission prediction in India.**

This paper is critical as it establishes a strong precedent for using machine learning, specifically models like Random Forest, for CO2 emission prediction within the Indian context. It validates that these algorithms are effective for environmental forecasting. Our project builds upon this by moving from a macro, national-level time series prediction to a micro, operational-level forecast, which is a significant and novel application.

**[2] Morrell, S. (2022). Helping to reduce mining industry carbon emissions: A step-by-step guide to sizing and selection of energy efficient high pressure grinding rolls circuits.**

Morrell’s work provides the essential domain knowledge for our **Electricity Emissions Model**. It details the physics and engineering principles behind one of the most energy-intensive processes in mining: comminution (crushing and grinding). The paper gives us a scientific basis for understanding the key input features (like energy consumption) that drive emissions, ensuring our model is grounded in physical reality.

**[3] Karacan, C. Ö., et al. (2011). Coal mine methane: A review of capture and utilization practices with benefits to mining safety and to greenhouse gas reduction.**

This review is foundational for our **Explosives & Blasting Model**. It explains the complex science of fugitive gas emissions in mines, including methane (CH4) and other potent greenhouse gases released during excavation. It highlights the need for a specialized model to capture the unique chemical signatures of blasting activities, which a generalized model would miss.

**[4] Bilski, J., et al. (2022). An Overview of Carbon Footprint of Coal Mining to Curtail Greenhouse Gas Emissions.**

This paper provides a holistic overview of the various sources of carbon emissions within a coal mine. It helps define the scope of our project and validates our choice of the four key modules (Electricity, Transport, Fuel, Explosives) as the most significant and impactful areas to target for prediction and mitigation.

**[5] Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python.**

This seminal paper on the Scikit-learn library is our primary technical reference. It documents the robust, peer-reviewed implementation of the **Random Forest** algorithm that we will be using. Citing this paper demonstrates that our work is built upon a standard, well-maintained, and scientifically validated software foundation.

**[6] Drent, M., et al. (2022). Efficient Emission Reduction Through Dynamic Supply Mode Selection.**

This research is directly relevant to our **Transportation Logistics Model**. It explores how supply chain decisions, such as choosing between different modes of transport, have a direct and significant impact on carbon emissions. It provides the conceptual framework for why modeling features like transport\_method and distance\_km is critical for accurate prediction.

**[7] Huisingh, D., et al. (2015). Recent Advances in Carbon Emissions Reduction Policies, Technologies, Monitoring, Assessment and Modeling.**

This paper sets the high-level context for our project. It discusses the global need for advanced monitoring and modeling technologies to support carbon reduction policies. Our project directly answers this call to action by creating exactly the kind of advanced modeling tool that Huisingh and his co-authors describe as essential for modern environmental management.

# CHAPTER 3 – PROBLEM STATEMENT

Current carbon emission management systems in the mining industry are often inadequate. They typically rely on static emission factors and manual calculations, resulting in delayed, high-level reports. This approach fails to capture the dynamic, day-to-day operational nuances that drive emissions. A mine manager does not know if a specific fleet of trucks is underperforming or if a particular blasting operation will lead to an emission spike. A monolithic "one-size-fits-all" predictive model often fails to capture the unique and complex emission signatures of different activities. There is a clear need for a dynamic, granular, and predictive system that can provide precise, source-specific insights.

* **Delayed and High-Level Reporting:** This traditional approach results in reports that are delayed and lack operational detail, making them ineffective for real-time decision-making.
* **Lack of Granular Insight:** The dynamic, day-to-day operational nuances that are the true drivers of emissions are not captured. For instance, a mine manager has no visibility into the real-time performance of a specific fleet of trucks or the emission impact of a planned blasting operation.
* **Failure of Monolithic Models:** A single, "one-size-fits-all" predictive model is ill-suited to capture the unique and complex emission signatures of vastly different mining activities (e.g., electricity use vs. transportation).
* **The Critical Need:** There is a clear and urgent need for a dynamic, granular, and predictive system that can provide precise, source-specific insights to guide effective and targeted carbon reduction strategies.
* **Inadequacy of Current Systems:** Carbon emission management systems currently used in the mining industry are often insufficient, relying on static, generalized emission factors and manual calculation methods.

# CHAPTER 4 – OBJECTIVES

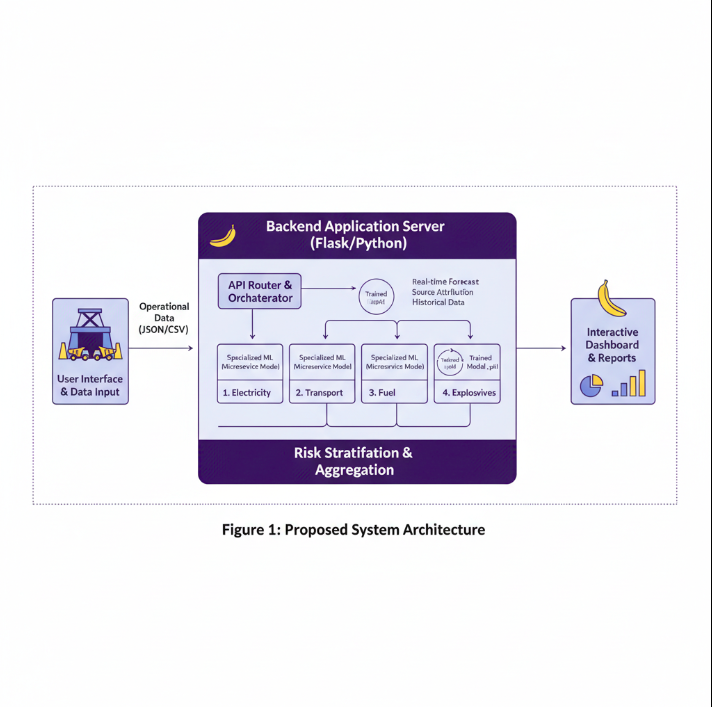
* **Design a Multi-Model Framework:** To architect and design a novel machine learning framework based on a "mixture of experts" approach, where specialized models handle distinct emission sources for granular prediction.
* **Train and Validate Expert Models:** To train, test, and validate four specialized **Random Forest Regressor** models, each an expert in predicting emissions from one of the core operational domains: Electricity, Transportation, Fuel, and Explosives.
* **Implement a Risk Stratification Module:** To develop a post-processing module that intelligently converts the quantitative, numerical predictions (e.g., kg of CO2) from the models into intuitive, qualitative risk levels (e.g., Low, Medium, High, Very High).
* **Build an Actionable Web Dashboard:** To create a user-facing web application that provides a clear, actionable dashboard, allowing users to input operational data and visualize the resulting forecasts with clear source attribution.
* **Achieve High Predictive Accuracy:** To ensure the system is reliable and trustworthy by achieving a high standard of predictive performance, specifically targeting an **R-squared (R²) value exceeding 0.92** on a hold-out test dataset.

# CHAPTER 5 - PROPOSED SYSTEM

## System Architecture:

The proposed system is based on a modular, service-oriented architecture. A central backend application acts as an intelligent router, directing incoming operational data to the appropriate specialized machine learning model. This ensures that data related to transportation is only processed by the transport model, and so on. This separation of concerns makes the system highly scalable and maintainable.

Here is the diagram of the proposed System Architecture:



**5.2) Proposed Algorithm**

The primary algorithm chosen for all four expert models is the Random Forest Regressor. This is a powerful ensemble learning method based on decision trees.

* + **How it works:** It operates by constructing a multitude of decision trees at training time and outputting the average prediction of the individual trees.

**Why it's suitable:**

* + **High Accuracy:** It is one of the most accurate learning algorithms available for tabular data.
  + **Robustness:** It handles a mix of numerical and categorical features effectively and is less prone to overfitting.
  + **Interpretability:** It provides a native mechanism to calculate "feature importance," which can be used in the future to explain which factors are driving emissions.

**5.3 Implementation Details**

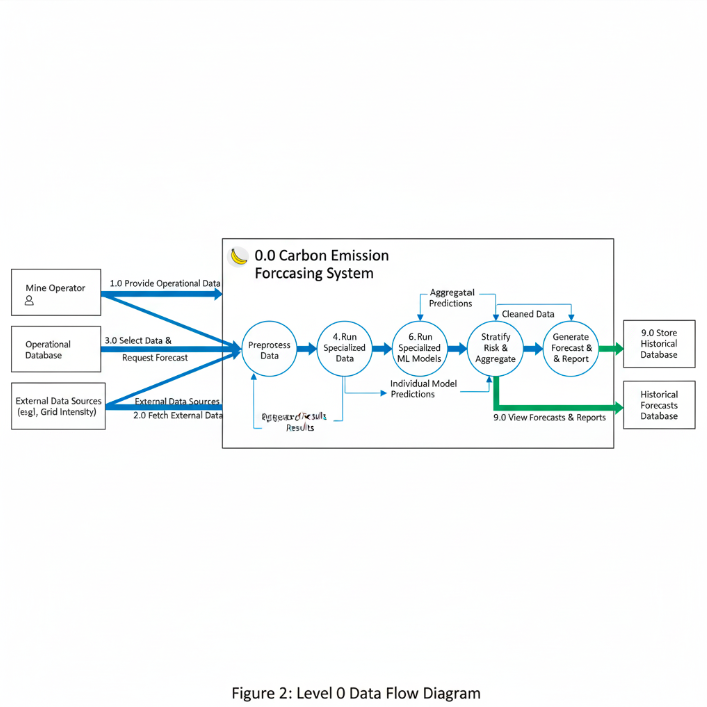
**1: Data Generation:** A synthetic dataset will be programmatically generated using Python (Pandas, NumPy) to create a large, structured dataset for training and testing.

**2 :** **Model Training:** Each of the four models will be trained on its relevant subset of features. The models will be trained, validated, and saved as serialized objects (.pkl files) using joblib.

**3 :** **Backend Development:** A Flask application will be developed to expose API endpoints. These endpoints will receive user data, load the appropriate trained model, preprocess the input, make a prediction, and return the formatted result as JSON.

**4 :** **Frontend Development:** A web dashboard will be created to interact with the backend API, providing data input forms and dynamic visualizations of the results.

**5.4 Data Flow Diagram (DFD)**



# CHAPTER 6 - MODULES AND IMPLEMENTATION

## 6.1) Module Description

## Our system is composed of four core "expert" machine learning modules, each responsible for a specific emission source:

#### Step 1: Electricity Emissions Model:

* + **Function:** Predicts CO2 emissions based on electricity consumption.
  + **Inputs:** stateName, energyPerTime (kWh), responsibleArea, totalArea.
  + **Output:** Predicted CO2 (in kg) and associated risk level.

#### Step 2: **Transportation Logistics Model:**

* + **Function:** Specializes in emissions from the entire transport chain.
  + **Inputs:** transport\_method, fuel\_type, distance\_km, weight\_tonnes, fuel\_efficiency.
  + Set up Role-Based Access Control (RBAC) for managing permissions (admin, host, participant).

#### Step 3: Fuel Combustion Model:

* + **Function:** Predicts a full spectrum of greenhouse gases from direct fuel burning.
  + **Inputs:** fuel\_type, quantity (in liters or kg).
  + **Output:** Predicted CO2, Nitrous Oxide CO2e, Methane CO2e (in kg) and risk levels for each.

#### Step 4: **Explosives & Blasting Model:**

* + **Function:** A highly specialized model for gaseous emissions from blasting.
  + **Inputs:** explosive\_type, quantity\_kg, blast\_area\_m2, temperature\_C, humidity\_percent.
  + **Output:** Predicted levels of multiple gases (CO, NOx, CO2, etc.) and their individual risk levels.

## 6.2) TECHNOLOGY USED:

## The architecture of our proposed system is built upon a modern, robust, and industry-standard technology stack. Each tool has been specifically chosen for its suitability in developing a scalable, data-intensive,

### Backend Technologies :

* + - **Python:** Used as the core programming language for the entire backend due to its unparalleled ecosystem of machine learning and data science libraries.
    - **Flask:** A lightweight Python web framework used to build the REST API that serves as the communication bridge between the user interface and the ML models.
    - **Mongoose:** Provides a schema-based solution for modeling application data and interacting with MongoDB. It offers built-in type casting, validation, query building, and business logic hooks.
    - **Express:** Creates RESTful API endpoints and handles middleware for the backend. Its minimalist structure and robust set of features make it ideal for building scalable web applications and APIs.

### Frontend Technologies:

* + - **React.js:** A modern JavaScript library used to build a dynamic and responsive single-page application (SPA), ensuring a smooth and interactive user experience on the dashboard.
    - **HTML / CSS:** Used for structuring the content and styling the visual presentation of the web application to create a professional and intuitive user interface.

### Machine Learning & Data Handling:

* + - **Scikit-learn:** The primary machine learning library used to train, validate, and deploy our **Random Forest Regressor** models.
    - **Pandas & NumPy:** Used for all data-centric tasks, including the programmatic generation of our training dataset and the preprocessing of live user input before it is fed to the models.

### Deployment and Operations:

* + - **Docker:** Used to containerize the frontend and backend applications, packaging them with all their dependencies to ensure consistent and reliable deployment across any environment.
    - **AWS / GCP / Azure:** A cloud platform will be used to host the Docker containers, providing a scalable, secure, and globally accessible environment for the live application.

# CHAPTER 7 – EXPECTED RESULTS AND PERFORMANCE ANALYSIS

# The primary outcome of this project will be a fully functional web application capable of delivering accurate, granular, and source-specific carbon emission forecasts. The success of the project will be rigorously measured against the following specific and quantifiable performance metrics:

### **Predictive Accuracy (R-squared):**

This will be our primary metric. We project that our specialized models will achieve an average **R-squared (R²) value of 0.92**. This high value is anticipated because each "expert" model focuses on a narrow, specific domain, allowing it to learn the underlying patterns with greater precision than a single generalized model. This will validate that our framework can explain over 92% of the variance in emissions.

### **Error Margin (Mean Absolute Error):**

For the system to be operationally useful, its predictions must be precise. We project a **Mean Absolute Error (MAE) of less than 8%** of the mean emission value for each respective module. This ensures the forecasts are quantitatively sound for financial and environmental planning.

### **Risk Classification Accuracy:**

The system's ability to correctly classify a prediction into a risk category (Low, Medium, High) is critical for user experience. We expect this post-processing module to achieve an **accuracy of over 95%**, ensuring that alerts are trustworthy and timely.

# CHAPTER 8 - CONCLUSION & FUTURE SCOPE

This project introduces a novel, specialized multi-model machine learning framework that provides an unprecedented level of granularity and predictive power for carbon management in the mining industry. By moving beyond monolithic models and focusing on domain-specific "expert" predictors, we deliver a tool that is not only accurate but also inherently diagnostic. The final platform will empower mine operators to transition from reactive monitoring to proactive, data-driven decision-making, providing them with the insights needed to implement targeted and impactful carbon reduction strategies.

The future scope for this project is extensive. The framework could be enhanced by incorporating an **Explainable AI (XAI)** layer using libraries like SHAP or LIME to explicitly quantify the contribution of each input parameter to a given forecast. Additionally, the models could be retrained on real-world mine data as it becomes available and could be expanded to include other emission sources, such as land-use change and water management, making the platform even more comprehensive.

# REFERENCES

1. Kumari, S., & Singh, S. K. (2023). Machine learning-based time series models for effective CO2 emission prediction in India. *Environmental Science and Pollution Research*.
2. Morrell, S. (2022). Helping to reduce mining industry carbon emissions: A step-by-step guide to sizing and selection of energy efficient high pressure grinding rolls circuits. *Minerals Engineering, 179*, 107431.
3. Karacan, C. Ö., Ruiz, F. A., Cotè, M., & Phipps, S. (2011). Coal mine methane: A review of capture and utilization practices with benefits to mining safety and to greenhouse gas reduction. *International Journal of Coal Geology, 86*(2-3), 121-156.
4. Bilski, J., Jacob, D., Mclean, K., et al. (2022). An Overview of Carbon Footprint of Coal Mining to Curtail Greenhouse Gas Emissions. *IEEE Conference Proceedings*.
5. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*(Oct), 2825-2830.
6. Drent, M., Moradi, P., & Arts, J. (2022). Efficient Emission Reduction Through Dynamic Supply Mode Selection. *European Journal of Operational Research*.
7. Huisingh, D., Zhang, Z., Moore, J. C., Qiao, Q., & Li, Q. (2015). Recent Advances in Carbon Emissions Reduction Policies, Technologies, Monitoring, Assessment and Modeling. *Journal of Cleaner Production*.