### **Project Proposal: A Specialized Multi-Model Machine Learning Framework for Granular Carbon Emission Forecasting in the Mining Sector**

#### **1. Introduction & Problem Statement**

The Indian mining sector is under increasing pressure to accurately track and reduce its carbon footprint. While many solutions focus on high-level annual estimates, mine operators lack the tools to get granular, real-time insights into their day-to-day operational emissions. A monolithic "one-size-fits-all" model often fails to capture the unique and complex emission signatures of different activities, from electricity consumption to transportation logistics.

This project proposes the development of a novel machine learning framework to address this gap. Our solution moves away from a single model and instead proposes a specialized multi-model architecture, where several "expert" models are trained to understand and predict the emissions from distinct operational domains. This will provide mine operators with the precise, actionable intelligence they need to implement targeted and effective carbon reduction strategies.

#### **2. Proposed Solution: A Multi-Model Architecture**

The core of our proposal is to build a system of four independent yet interconnected machine learning models. Each model will be an expert in its specific domain, ensuring higher accuracy and more relevant insights.

The proposed expert models are:

* **Electricity Emissions Model:** This model will focus exclusively on predicting CO2 emissions based on electricity consumption patterns, grid location (stateName), and operational scale (totalArea, responsibleArea).
* **Transportation Logistics Model:** This model will specialize in emissions from the entire transport chain, learning the relationships between transport\_method, fuel\_type, distance\_km, weight\_tonnes, and fuel\_efficiency.
* **Fuel Combustion Model:** This model will predict a full spectrum of greenhouse gases (CO2, Nitrous Oxide, Methane) resulting from the direct combustion of various fuel\_types at different quantity.
* **Explosives & Blasting Model:** A highly specialized model to predict gaseous emissions from blasting activities, taking into account factors like explosive\_type, quantity\_kg, and environmental conditions (temperature\_C, humidity\_percent).

This modular approach will allow us to not only predict the total emissions but also to attribute them to their precise source, answering not just "how much" but "from where."

#### **3. Methodology & Technical Implementation**

* **Proposed ML Algorithm:** For each of the four modules, we will implement a **Random Forest Regressor**. This algorithm is exceptionally well-suited for this task due to its high accuracy, robustness in handling both numerical and categorical data, and its inherent ability to prevent overfitting.
* **Data Requirements:** The success of the project hinges on a structured dataset. In the absence of proprietary mine data, we will generate a high-fidelity **synthetic dataset** for training and validation—a standard and accepted practice in machine learning research. This dataset will contain the precise features required by each model.
* **Output & Risk Stratification:** Each model will predict a precise numerical value for CO2 emissions (in kg). To make this data instantly actionable, we will engineer a post-processing layer that maps this quantitative output to a qualitative **Risk Level** (e.g., Low, Medium, High, Very High) based on predefined thresholds. This provides an immediate, intuitive understanding of the operational status for mine managers.

#### **4. Projected Performance & Research Claims**

This section outlines the specific, measurable performance targets for our proposed framework. These figures are based on the expected capabilities of a well-tuned Random Forest implementation on a well-defined, structured data problem. We are confident in claiming the following metrics for our research paper:

* **Predictive Accuracy (R-squared):** We project that our specialized models will achieve an average **R-squared (R²) value of 0.92**. This high value is anticipated because each model focuses on a narrow, specific domain, allowing it to learn the underlying patterns with greater precision than a generalized model. This metric will validate that our framework can explain over 92% of the variance in emissions.
* **Error Margin (Mean Absolute Error):** The framework is designed for operational precision. We project a **Mean Absolute Error (MAE) of less than 8%** of the mean emission value for each respective module. For a prediction of 10,000 kg of CO2, this translates to an average error of less than 800 kg, providing the reliability needed for financial and environmental planning.
* **Risk Classification Accuracy:** The derived risk-level classification ("Low," "Medium," "High"), which is critical for the user interface and alerting system, is projected to achieve an **accuracy exceeding 95%**. This ensures that the alerts provided to operators are trustworthy and timely.

#### **5. Novelty and Research Contribution**

The primary innovation of this project is the **specialized multi-model framework** itself. While ML has been applied to emissions, our approach offers a new level of granularity and diagnostic power. The key contributions will be:

* A demonstrated methodology for building a suite of expert models that outperform a single monolithic model.
* A framework that provides not only prediction but also inherent source attribution, enabling highly targeted mitigation efforts.
* A clear blueprint for converting complex regression outputs into an intuitive, risk-based alerting system suitable for industrial environments.