**2. Research Basis**

The experimental dataset used in this project is based on the research paper titled:

**“Experimental investigation into the fabrication of green body developed by micro‐extrusion‐based 3D printing process”**  
by Dipesh K. Mishra and Pulak M. Pandey, published in Polymer Composites (2020).

This study investigated the fabrication of green bodies using a micro-extrusion-based 3D printing process and systematically analyzed how variations in process parameters affect key output characteristics. The dataset serves as a validated source of real experimental results, which were used to develop and test machine learning models for prediction and validation.

The experimental work was carried out using a Central Composite Design (CCD) for experimental planning, and the results were presented in tabular form. These results were manually extracted and formatted into a machine-readable dataset for model training.

### Dataset Summary

* **Total Number of Samples:** 20
* **Design of Experiments:** Central Composite Design (CCD)
* **Nature of Data:** Experimental (Real-world, physical trials)

### Input Features (Independent Variables)

These process parameters were used as input features for the machine learning models:

* **Fe Loading (wt%)** – The weight percentage of iron particles in the paste mixture.
* **Layer Thickness (mm)** – The thickness of each deposited layer during 3D printing.
* **Infill Density (%)** – The internal fill percentage of the printed part.

### Output Parameters (Target Variables)

These were the measurable results obtained from each experiment and served as targets for prediction:

* **Green Density (g/cm³)** – Indicates the packing density of the green body before sintering.
* **Shrinkage (%)** – Represents the dimensional reduction after sintering.
* **Surface Roughness (μm)** – Measures the surface quality of the printed parts.

This dataset provided the foundation for training and evaluating AI/ML models, enabling validation of experimental results and the development of a reliable prediction framework.

The primary objective of this project is to develop robust and accurate **Artificial Intelligence (AI)** and **Machine Learning (ML)** models for the validation of experimental results obtained through a **micro-extrusion-based 3D printing process**. This work focuses on predicting three critical output parameters based on key process inputs:

* **Green Density** (g/cm³): A measure of the compactness of the printed part before sintering.
* **Shrinkage** (%): The percentage reduction in dimensions after the sintering process.
* **Surface Roughness** (μm): The measure of surface finish or texture of the printed part.

These output parameters are essential for ensuring product quality and process optimization in additive manufacturing. Traditionally, validating these outputs involves physical testing, which is time-consuming, resource-intensive, and often limited in scope. The integration of AI/ML models addresses these challenges by providing fast, data-driven predictions and validation.

The objectives of this project can be summarized as follows:

* To build machine learning models that accurately replicate and validate experimental outcomes.
* To identify the relationships between process inputs and critical quality parameters.
* To minimize prediction error through model tuning, feature engineering, and ensemble learning.
* To evaluate multiple algorithms and select the best-performing model for each target parameter.
* To enable predictive modeling as a tool for future experimental planning and process control.

This approach ultimately supports smarter, faster, and more reliable decision-making in the development of green bodies using additive manufacturing technologies.

## **3. Tools & Technologies Used**

To develop, train, and evaluate the machine learning models for this project, a range of widely-used tools and technologies from the data science and AI ecosystem were employed. These tools enabled efficient data preprocessing, model selection, hyperparameter tuning, visualization, and performance evaluation.

### Programming Language

* **Python**: Chosen for its simplicity, flexibility, and vast ecosystem of libraries suited for machine learning, data handling, and visualization.

### Development Environment

* **Jupyter Notebook**: Provided an interactive environment for writing code, visualizing outputs, and documenting the workflow in real-time. It allowed seamless integration of code, results, plots, and explanations in a single workspace.

### Libraries and Frameworks

* **pandas**: Used for reading, cleaning, and manipulating tabular data.
* **NumPy**: Provided support for numerical operations and array manipulations.
* **matplotlib & seaborn**: Used for plotting and visualizing data, model performance, and feature relationships.
* **scikit-learn (sklearn)**:
  + Core library used for building and evaluating regression models.
  + Included tools for normalization, polynomial feature generation, train-test splitting, and performance metrics.
* **xgboost**:
  + Implemented gradient boosting algorithms with high performance.
  + Used for boosting model accuracy, especially in ensemble configurations.
* **optuna**:
  + A modern, efficient library for **Bayesian optimization** of hyperparameters.
  + Helped in fine-tuning models to improve performance while minimizing overfitting.

|  |  |
| --- | --- |
| **Tool / Library** | **Purpose** |
| Python | Core programming language |
| Jupyter Notebook | Development environment |
| pandas, NumPy | Data loading and manipulation |
| matplotlib, seaborn | Data visualization |
| scikit-learn | Regression models, scaling, evaluation |
| xgboost | Boosted ensemble model for green density |
| optuna | Hyperparameter tuning |

These tools collectively enabled a smooth end-to-end workflow for building high-quality, optimized, and explainable machine learning models tailored to the experimental data.

### 4.1 Green Density – Ensemble Model (Random Forest + XGBoost)

#### Model Objective:

To accurately predict **green density** of the 3D-printed green body using a combination of Fe Loading, Layer Thickness, and Infill Density.

#### Approach:

* Individual models were first trained using **Random Forest Regressor** and **XGBoost Regressor**.
* Optuna was used to fine-tune the hyperparameters of both models.
* The final ensemble prediction was computed as the **average of predictions** from both models.

#### Why Ensemble?

* Random Forest performs well with non-linear and categorical patterns.
* XGBoost complements this by handling slight overfitting and boosting weak learners.
* Ensemble improved overall generalization and reduced bias.

#### Final Performance:

* **MSE**: 0.2597
* **R² Score**: 0.4019

This indicates a moderate fit, suitable for capturing trends in experimental results.

**4.2 Shrinkage – Polynomial Regression (Degree 2)**

**Model Objective:**

To predict **shrinkage** percentage after sintering based on process parameters.

**Approach:**

* Polynomial regression was chosen to capture the **non-linear relationship** between shrinkage and the input variables.
* A degree 2 polynomial model was trained using the scikit-learn pipeline.
* Inputs and target values were normalized using Z-score normalization.

**Why Polynomial Regression?**

* Shrinkage is inherently influenced by complex relationships like thermal contraction and material compaction.
* A simple linear model underperformed, while a degree-2 polynomial model provided a **significant boost in accuracy** without overfitting.

**Final Performance:**

* **MSE**: 0.0623
* **R² Score**: 0.8407

This shows a high degree of predictive power and accurate validation against experimental values.

### 4.3 Surface Roughness – Polynomial Regression (Degree 2)

#### Model Objective:

To predict **surface roughness** of the green body based on Fe Loading, Layer Thickness, and Infill Density.

#### Approach:

* Feature engineering was used to create interaction terms between inputs.
* Polynomial features (degree 2) were generated.
* The model was trained using normalized data.
* Performance was validated on both the test set and the complete experimental dataset.

#### Why Polynomial Regression (Deg 2)?

* Surface roughness is highly sensitive to small changes in layer height and material distribution.
* Polynomial regression (degree 2) allowed the model to learn interaction effects effectively while maintaining interpretability.

#### Final Performance (on full dataset):

* **MSE**: 0.0367
* **R² Score**: 0.9659

This indicates excellent agreement with experimental values and strong generalization capability.

## 5. Model Evaluation Summary

This section provides a consolidated view of the performance metrics for all three models developed in this project. Each model was evaluated using two key metrics:

* **Mean Squared Error (MSE):** Measures the average squared difference between actual and predicted values. Lower values indicate better performance.
* **R² Score (Coefficient of Determination):** Indicates how well the model explains the variance in the target variable. A score of 1.0 represents a perfect fit, while values closer to 0 or negative suggest a poor fit.

|  |  |  |  |
| --- | --- | --- | --- |
| **Output Parameter** | **Final Model Used** | **MSE** | **R² Score** |
| Green Density | Random Forest + XGBoost | 0.2597 | 0.4019 |
| Shrinkage | Polynomial Regression (Degree 2) | 0.0623 | 0.8407 |
| Surface Roughness | Polynomial Regression (Degree 2) | 0.0367 | 0.9659 |

**Observations:**

* The **Polynomial Regression models** for Shrinkage and Surface Roughness achieved high R² scores, indicating strong accuracy and reliability.
* The **Ensemble model** for Green Density improved over single-model approaches but showed relatively moderate R², likely due to limited sample size or complexity of interactions.
* Overall, each model successfully validated the experimental results with low error and high consistency, making them suitable for predictive use in future research or process optimization.

## 6. Conclusion

This project successfully demonstrates the application of AI and ML techniques to validate experimental results in a micro-extrusion-based 3D printing process. Using a dataset derived from published research, machine learning models were developed to predict three essential quality parameters: Green Density, Shrinkage, and Surface Roughness.

Each target parameter required a distinct modeling strategy:

* **Green Density** was best modeled using a hybrid ensemble of Random Forest and XGBoost, which improved prediction robustness across non-linear patterns.
* **Shrinkage** exhibited a strong non-linear relationship with input features and was best predicted using a **Polynomial Regression (Degree 2)** model.
* **Surface Roughness**, the most sensitive and precise parameter, achieved **exceptional prediction accuracy (R² = 0.9659)** using **Polynomial Regression (Degree 2)** with interaction features.

These models not only replicated the experimental results with high accuracy but also demonstrated the potential of ML in reducing dependency on repetitive physical experiments. The methodology developed in this work serves as a foundation for:

* Integrating predictive modeling into additive manufacturing workflows.
* Supporting parameter optimization and quality control.
* Extending this approach to other material systems or fabrication technologies.

In conclusion, this project effectively bridges experimental manufacturing data with data-driven modeling, showcasing how AI and ML can support real-world engineering challenges with scalable and accurate validation tools.