**Analyzing Chemical Compounds Using Machine Learning**

**Overview**

This project leverages **Python** and **Machine Learning** to explore correlations between the **refractive index** of chemical compounds and various physical properties. The goal is to determine how different attributes like **melting point, boiling point, dipole moment, and relative density** impact the refractive index.

**Dataset**

* A compiled CSV file containing **42 selected chemical compounds**.
* Parameters studied:
  + **Melting Point**
  + **Boiling Point**
  + **Dipole Moment**
  + **Refractive Index**
  + **Epsilon** (Dielectric Constant)
  + **Relative Density (Rd)**
  + **Mu (Chemical Potential)**
  + **Polarity**

**Methodology**

1. **Data Preprocessing**:
   * Imported dataset using **Pandas** and performed data cleaning.
   * Converted numerical variables to appropriate **float** formats.
   * Checked dataset consistency using df.describe() and df.dtypes().
2. **Exploratory Data Analysis (EDA)**:
   * Generated **pair plots** and **correlation heatmaps** using **Seaborn**.
   * Visualized relationships between **refractive index** and other physical properties.
   * Identified a **strong correlation between refractive index and relative density (Rd)**.
3. **Machine Learning & Regression Modeling**:
   * Implemented a **linear regression model** using a custom **gradient descent function**.
   * Calculated **correlation coefficients** for different variables.
   * Determined that **relative density (Rd) has the strongest impact on refractive index**.
4. **Data Visualization**:
   * Plotted scatter plots to analyze correlations between refractive index and properties like **dipole moment, boiling point, and epsilon**.
   * Used **point plots and categorical plots** to analyze variations in refractive index across different compounds.

**Key Findings**

* **Refractive index has the strongest correlation with relative density (0.60).**
* Other properties such as **dipole moment and epsilon** show minimal correlation.
* The study demonstrates how **Python-based data analysis and visualization techniques** can be used to evaluate chemical compound properties effectively.

**Tools & Libraries Used**

* **Python Libraries:** Pandas, NumPy, Seaborn, Matplotlib
* **Machine Learning Techniques:** Linear Regression, Correlation Analysis
* **Data Processing & Visualization:** Heatmaps, Scatter Plots, Pair Plots, Regression Models

**Conclusion**

This project successfully explored correlations between the refractive index and other chemical properties. The findings show that **relative density** plays a significant role in determining refractive index, while other features have a weaker impact. Future work can explore additional **statistical models and advanced machine learning techniques** to refine these insights further.

For more details, check the **notebook file** or reach out via **GitHub**.