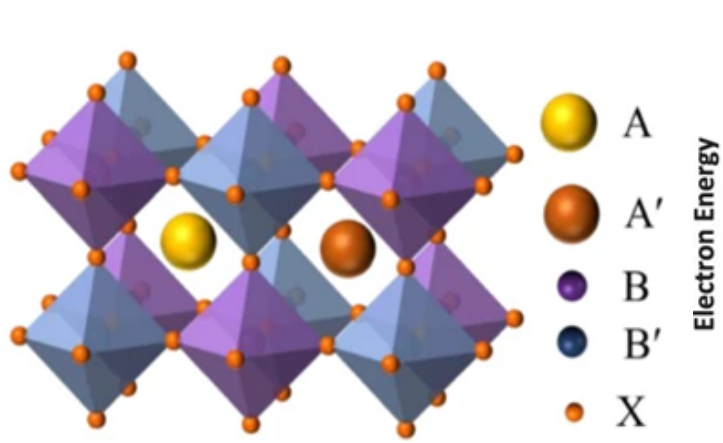
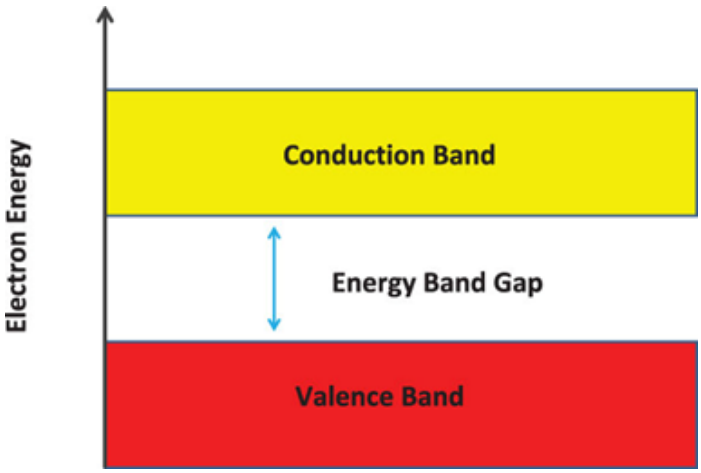
EXCAVATE

Problem Statement

Unveiling Band Gaps in Perovskite Oxides for Next-Gen Electronics

Problem:





Perovskite oxides are a class of materials with **tunable electronic** properties, making them valuable for **solar cells, LEDs, semiconductors, and optoelectronic devices**. A key property that determines their usefulness is the **band gap (Eg)**, which defines whether a material is an **insulator, semiconductor, or conductor**. We aim to use **machine learning (ML) models** to predict band gaps quickly and efficiently.
Your goal is to **analyze, preprocess, and train ML models** using a real-world **materials science dataset** and compete to achieve the most **accurate predictions**.

Participants will tackle two machine learning tasks:

- **Binary Classification Task (MC - Model 1)**
  Classify perovskite materials as **insulators (Eg ≥ 0.5 eV)** or **non-insulators (Eg < 0.5 eV)**.
- **Regression Task (MR - Model 2)**
  Predict the **exact band gap (Eg in eV) for insulators only**.

The dataset consists of **5,152 perovskite compositions**, with **structural and atomic features**. Each row represents a **unique perovskite material**, and columns describe its **composition, atomic properties, geometric stability, and electronic properties**.

Compositional Features:

Column Name	Physical Meaning
functional group	The perovskite chemical formula (e.g., $AgBaAuCdO_6$).
A, A_OS	A-site cation and its oxidation state.
A', A'_OS	Second A-site cation (for double perovskites) and its oxidation state.
B, B_OS	B-site cation and its oxidation state.
B', B'_OS	Second B-site cation (for double perovskites) and its oxidation state.

Electronic (Atomic) Descriptors:

Column Name	Physical Meaning
HOMO, LUMO	Highest and Lowest Molecular Orbital energy levels, affecting electron transfer.
Ionization Energy (IE)	Energy required to remove an electron. Higher values indicate insulating behavior.
Electronegativity (X)	Tendency of an atom to attract electrons, affecting band gap.
Electron Affinity (EA)	Energy released when an atom gains an electron.
Zunger's Pseudopotential Radius (Z-radius)	Effective atomic size influencing bonding.

Structural (Geometric) Descriptors:

Column Name	Physical Meaning
Tolerance Factor (t)	Measures perovskite structural stability.
Octahedral Factor (μ)	Measures B-site cation size compatibility in the structure.
Mismatch Factor (A, B sites)	Size differences between cations affecting stability.

Target Variables:

Column Name	Physical Meaning
PBE Band Gap (Eg in eV)	Energy gap between valence and conduction bands.

Note:

- Use **80% - 20%** train-test split for the dataset.
- Create **classification model/models** to classify the materials into **insulators (PBE band gap > 0.5 eV)** and **non-insulators (PBE band gap < 0.5 eV)**. Use all the data for classification.
- Create model/models to predict the **band-gap values**. Use only the materials identified as insulators (PBE band gap > 0.5 eV).
- Use data analytics tools to extract insights into the dataset.
- Perform **post processing analysis** of the model to provide insights into the role of features.

Submission Criteria:

- Participants need to submit a .zip file consisting of the **(.ipynb) file** of the **solution along with the brief PDF Report (maximum 5 pages)**.
- Use **Python** and **Google Colaboratory** to write your code.
- Complete **each task in a separate section** and **add comments** to your code wherever necessary.
- Use the standard packages for data processing and regression. Clearly **document any external packages** used by your code.
- Submission is to be done only on Email: **excavate.composit@gmail.com**
- Please don't forget to mention your **name, College Name, Composit ID** and **Team ID** in the mail.

Contact :

Suryansh Singh : 9670006875
Aarya Amit Joshi : 7045020804