SPECTRUM DECODERS

INTRODUCTION

- Infrared (IR) Spectroscopy is crucial for molecular analysis, identifying vibrational modes and chemical interactions.
- Traditional quantum mechanical methods like Density Functional Theory (DFT) are accurate but computationally expensive and time consuming.
- Machine Learning (ML) provides an alternative for predicting IR spectra efficiently while maintaining accuracy.

SMART AI ACADEMY

Interdisciplinary AI Applications

- Shows how AI can accelerate research and reduce computational costs in high-performance computing fields.

Applications in AI Research & Development

- Demonstrates multi-output regression models and clustering for spectral prediction.
- Uses TensorFlow and Random Forest models, relevant for AI and data science training programs.

AI-Driven Scientific Advancements

- The project integrates AI and quantum chemistry, making it a perfect fit for an AI-focused institution.
- Uses Machine Learning (ML) to solve real-world challenges in molecular spectroscopy and material science.

Educational & Training Potential

- Can be integrated into AI certification programs to teach real-world AI applications in science field.
- The project can be used as a case study in AI courses covering:
- Supervised learning (Random Forest, Multi-Output Regressors)
- Feature engineering and dataset optimization
- AI-driven scientific research and computational modeling

AI for Sustainability & Industry Applications

- AI-driven spectral analysis can optimize industrial processes (e.g., drug designing).

Machine Learning Techniques Used

MULTI OUTPUT REGRESSOR

- Handles multiple target variables simultaneously for predicting vibrational frequencies and intensities.
- Enables efficient and indepentedent prediction of each target variables.

RANDOM FOREST REGRESSOR

- An ensemble method that builds multiple decision trees to enhance predictive accuracy and reduce overfitting.

Model Architecture and Design

Data Collection

- Uses datasets from Gaussian 16 software with molecular geometries, vibrational frequencies, and quantum properties.

Data Preprocessing

- Includes handling missing values, feature scaling, and label encoding.

Feature Engineering

- Extracts molecular descriptors such as atomic coordinates, bond lengths, dipole moments, and symmetry.

Model Training

- Uses Random Forest with Multi-Output Regressor to predict IR spectra.

Prediction & Visualization

- Generates IR spectra using Gaussian curves and compares actual vs. predicted results.

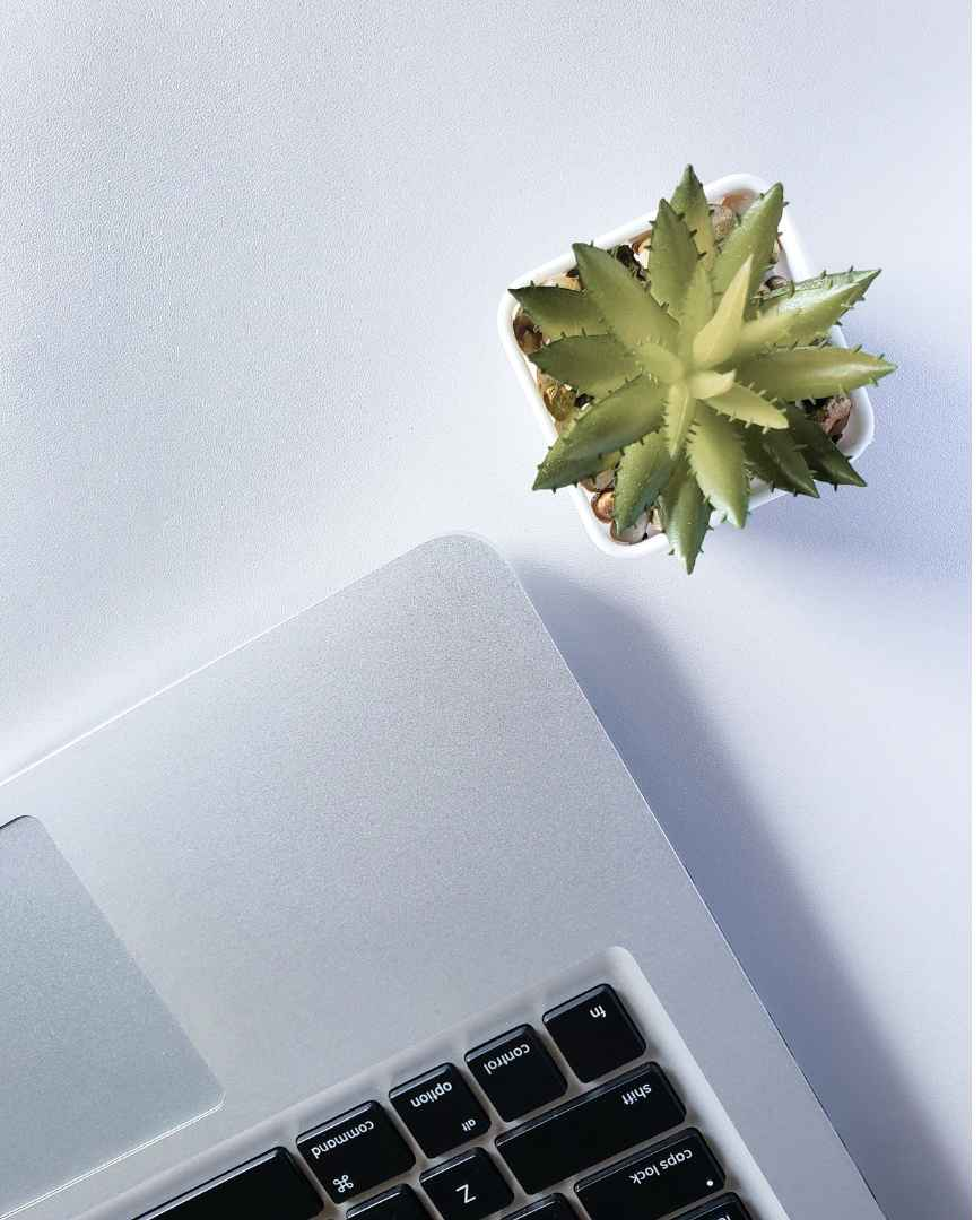
RESULTS

Model Performance:

- Smaller dataset: High accuracy ($R^2 = 0.90$, MAE = 7.53 cm^{-1})
- Larger dataset: Accuracy drops ($R^2 = 0.65$) due to increased complexity.

Clustering of Molecules:

- Five clusters identified based on vibrational and structural properties.
- Clusters include simple hydrocarbons, halogenated molecules, and inorganic compounds.



FUTURE SCOPE:

- Larger and More Diverse Datasets
- National Biomedical Resource for Advanced ESR Spectroscopy

Sustainable Development Goals

GOAL 4: QUALITY EDUCATION

- This project can serve as an educational tool for teaching AI application in chemistry.

GOAL 9: INDUSTRY,INNOVATION & INFRASTRUCTURE

- Machine learning streamlines materials discovery,enhancing industries like chemical engineering, nanotechnology, and material sciences.
- Reduces reliance on high-performance computing (HPC), making scientific computations more accessible.

CONCLUSION

- ML significantly enhances IR spectral predictions, offering a cost-effective and scalable alternative to quantum chemistry methods.
- The integration of ML models with molecular spectroscopy paves the way for faster and more accurate molecular characterization.