Riya Elizabeth John

GitHub: [Riyabelle25]
LinkedIn: [riya-elizabeth-john-3b57201ab]
Email: riyaelizabethj@gmail.com
Time zone: UTC +5:30
Town, country: Surat, India

Machine Learning Model for the Albedo of Mercury

Meeting with Mentors:

- **Before May 30th:** Reachable anytime between 06:30 am to 6:30 pm (UTC) [12:00 pm to 12:00 am IST] through Email/any other convenient means
- **After May 30th:** Reachable anytime between 3:30 am to 8:30 pm (UTC) [9:00 am to 2:00 am IST] through Email/any other convenient means

Can join a planned video session as well.

OVERVIEW

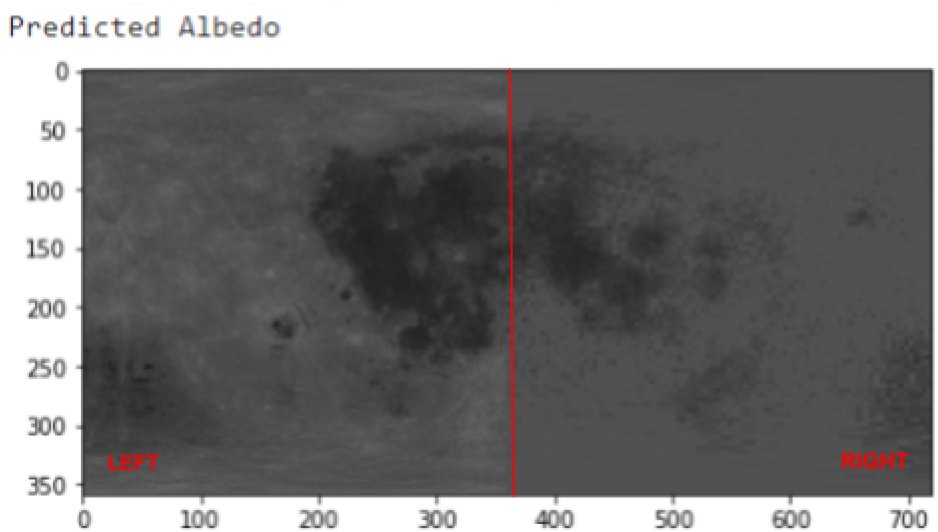
I aim to identify the hidden relationships between given planetary mapped datasets of chemical composition and albedo, using Machine Learning techniques-with the goal of gaining deeper understanding of planetary surfaces and thus predict chemical composition for planetary surfaces with incomplete datasets.

PROBLEM DESCRIPTION

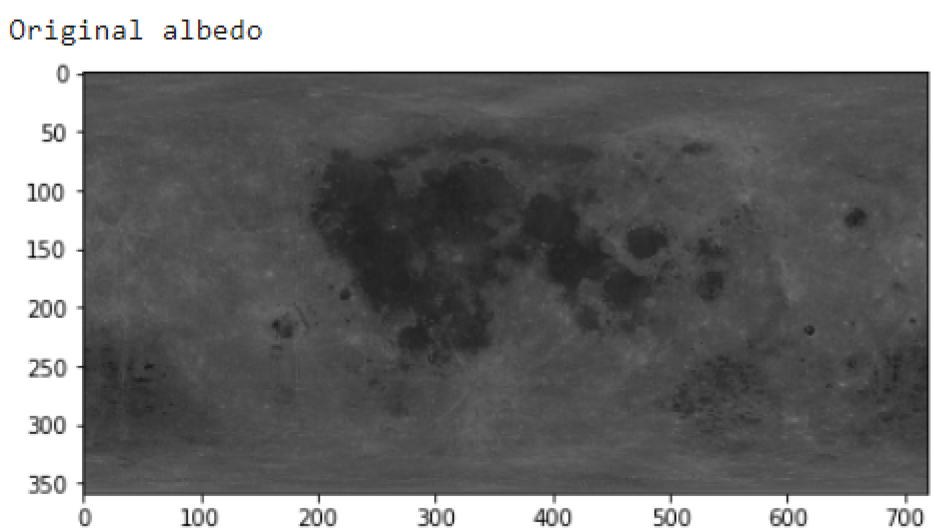
Planetary surfaces are observed through a wide range of the electromagnetic spectrum (e.g. radar, infrared, optical, ultraviolet, x-ray, gamma-ray), and each wavelength provides unique information about the chemistry, mineralogy, and history of the surface. Yet the information is not entirely independent. For example, the chemical element iron, which is mapped with x-rays and gamma rays, is highly related to optical albedo on the Moon.

I have confirmed this correlation, in my evaluation task:

The left half of the below map is the Actual Albedo, and the right half is the Predicted Albedo. I used **GradientBoostingRegressor** to map the chemical composition of the Moon's surface to the albedo.



This below, is the complete map of the Original Albedo. As you can see, the model predictions are highly accurate.



Also, I got an extremely **low Mean Square Error of 0.000940**, which further corroborates the correlation.

Observing this correlation, we can develop high-spatial resolution predictive maps of iron based on optical data. On other planets however, the relationships between the observations are less well known, and some datasets are missing- thus making the predictive powers of Machine Learning modelling, a valuable tool for analysis.

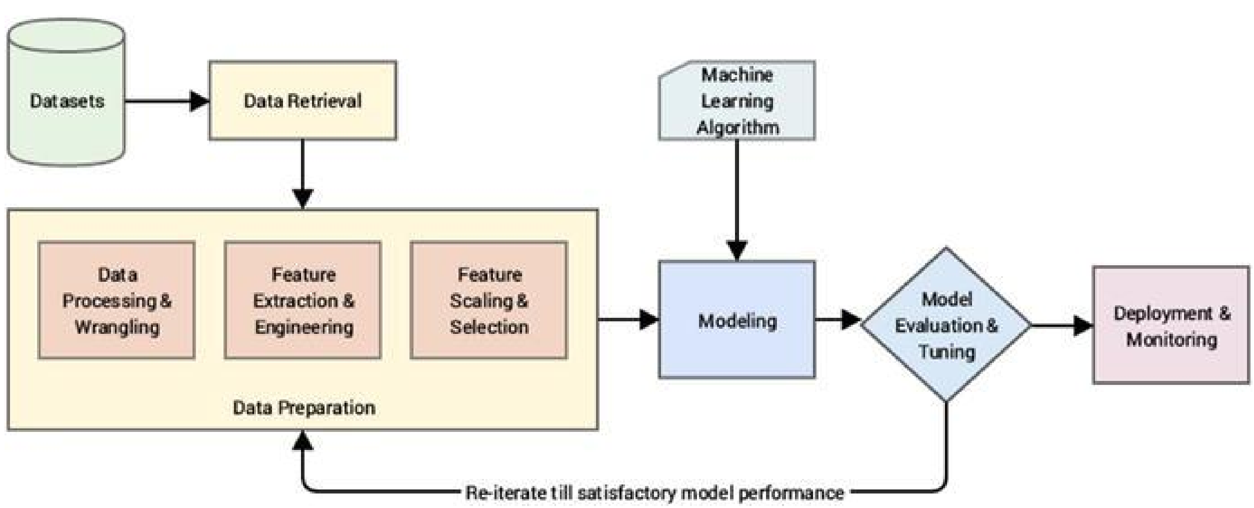
SOLUTION

We seek to study planetary surfaces by using maps of surfaces at available wavelengths to discover the relationships between the measurements and to make predictions about chemical composition. This provides a way of studying the geologic history of a planet with existing data, which is valuable given the infrequent opportunities for new measurements by planetary spacecraft.

GOALS:

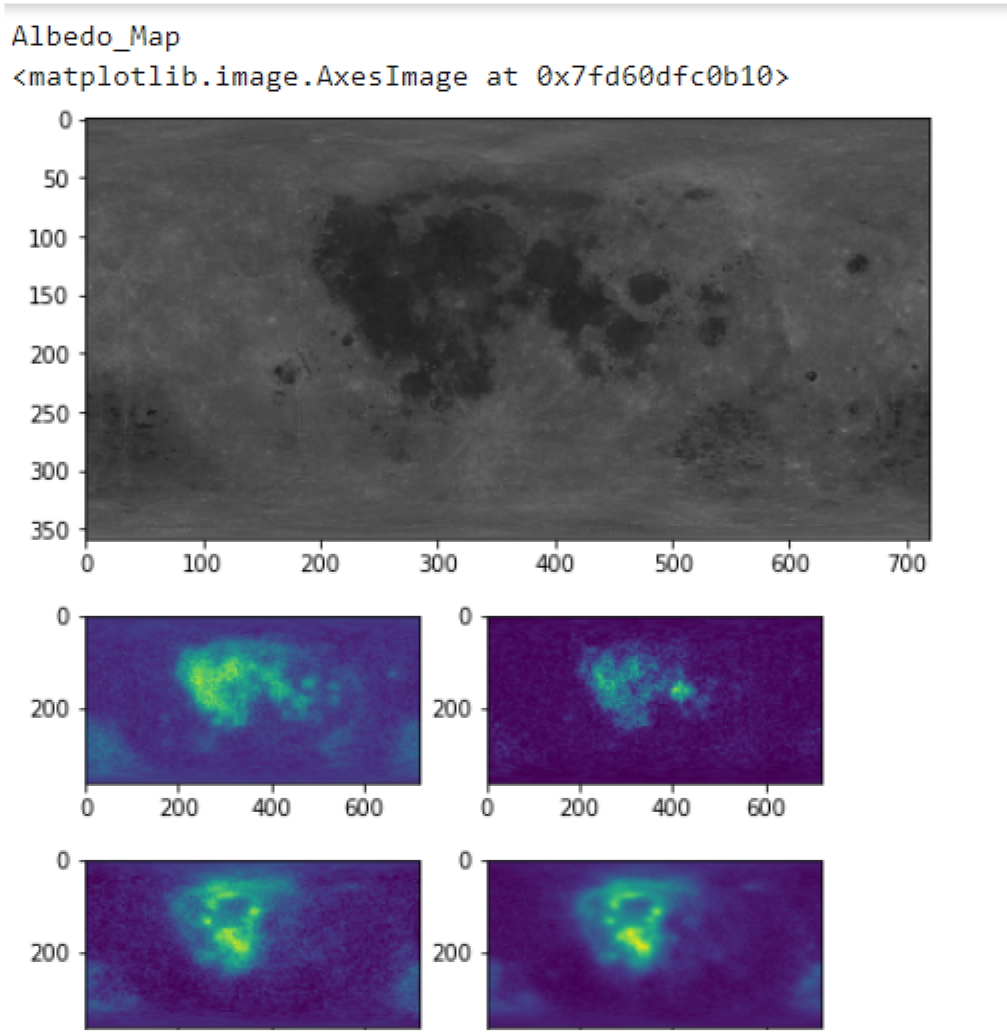
1. Selection of candidate machine learning techniques, including a Convolutional Neural Network architectures well suited to the problem.
2. Modelling of Mercury albedo and chemical composition based on Messenger mission data
3. Identify evaluation criteria for each technique, and hyper-tune the models
4. Discuss results and implications for planetary science; write a report with neat inline documentation

IMPLEMENTATION PLAN:



Phase 1: Data Exploration and Preparation (June 07-June 21)

- Having already performed preliminary Data Analysis for my Evaluation Task, I would dive deep with EDA (Exploratory Data Analysis) - employ better techniques more suitable for studying planetary mapped data. Study [this](#) paper to better understand data.
  Currently I’m using sklearn and seaborn for visualization:



- After EDA, I’ll go ahead with data preparation, especially check for outliers, and normalize the data.

Phase 2: Modelling (June 21- July 20)

For modelling, I will perform experiments based on two approaches:

1. **ML Regression Models with Multiple Target Variables:** This paper has a good overview of the model approaches to multi-target regression. I aim to spend two weeks implementing these, and note my observations simultaneously.
2. **Regression with Convolutional Neural Networks in PyTorch:** CNNs have been shown to be a good fit for making predictions from imaging data. Here, I shall build a CNN architecture using Pytorch for regression analysis and record my observations. This task should also take not more than 2 weeks.

Phase 3 Model Evaluation/Tuning (July 20- July 31)

- Carry out evaluation and hyper-tuning of the models (after deciding a performance metric).
- Discuss the outcome of modelling approaches with the mentors.
- Write a comprehensive report on all the experiments, and identify the best performing model.

Phase 4 (August 01- August 16)

Buffer period in case of unavoidable delays. I will make use of this time for code cleanup, and polishing inline documentation.

TIMELINE:

April 10- May 17 (Homework period)	Study planetary composition and imaging data mapping, look out for better ML techniques, Neural Networks for Regression
May 17- June 07 (Bonding Period)	Discuss the tasks ahead, learn more about the Messenger data from the mentors
June 07- June 14	Exploratory Data Analysis
June 15- June 21	Data Preprocessing/Preparation
June 21- July 04	**Modelling Approach 1**: Regression with Multiple Target Variables
June 05- July 19	**Modelling Approach 2**: Regression with CNNs in Pytorch
July 12- July 16	**Evaluations, implement suggestions**

July 20- 31 July	Model Evaluation and Tuning, Report writing
August 01- August 16	Code clean-up, and complete pending tasks, if any
August 16 - August 23	**Final Evaluations, implement suggestions**
August 31	**Results Announced!**

ABOUT ME

I'm an Undergraduate Sophomore at the Indian Institute of Technology, Roorkee, India.
My areas of interest lie in Python programming, and applying machine-learning/ deep-learning methodologies to solve interesting research problem statements and build cool software applications.
I enjoy participating in the Tech community of my college and am both:

- An active student developer at the [Mobile Development Group](), one of the tech groups of my college (we foster the Open Source Software Development culture in the campus) , and
- A Core member of the [Vision and Language Group]() (we foster a research-centric Deep Learning Community on the campus).

My activities have led me to collaborate and contribute to several Open Source projects here, hence I'm fluent with version control softwares like Git.

MOTIVATION:

Being a student developer with a passion for research and the field of machine learning, I appreciate the significance of FOSS and the ML4Sci organization caught my eye.
Ever since my debut as a novice developer, it has always been a dream to contribute to open source projects- especially when there's a good, active community behind it!
Besides developing applications, I also have a keen interest in research, especially Deep Learning. Engaging with ML4Sci will thus help me whet my interest in both.

WHY ME?

In order to upscale my Machine Learning (and Python) skills, I completed certified courses on Coursera and have done projects on the same as well- increasing my familiarity with major ML libraries like sklearn, pandas, Keras, Tensorflow/Pytorch etc.

Being a part of the Vision and Language Group, IIT Roorkee, I am also familiar with major Deep Learning methodologies- we regularly read and discuss various research papers from top tier conferences.

Recently, my internship with The Sparks Foundation introduced me to DL based time series forecasting. I worked on Stock Market Prediction using Numerical and Textual Analysis; I carried out Multivariate Time Series Forecasting using Keras' LSTM. My work is here.
(Also, all my projects and certificate links are mentioned in my CV)

Most importantly, I love learning- I'm a fast learner, and believe that hard work trumps talent. From my experiences so far, I know that I am ready to expand my skills yet again.

I can happily devote **30-35** hours a week, as I don't have other internships or commitments this Summer.

MY OTHER COMMITMENTS:

May 17- 31: My End-Semester exams are scheduled at this time, but I will be available via Email/Gitter/ or any other convenient means for discussions and other communication.

Asides from these, I have no other pressing commitments this Summer, and will keep the community and mentors posted for updates.

REFERENCES

1. https://www.sciencedirect.com/science/article/pii/S001910352030107X
2. http://cig.fi.upm.es/articles/2015/Borchani-2015-WDMKD.pdf
3. My Evaluation Task: https://colab.research.google.com/drive/1NeMoRsdqi9I2OBOuPawQQatue8T82oAq?usp=sharing