Predicting the Influence of Environmental Factors on Green Bond Yields

Team AlgoRythms: Avani Sharma, Shalagha Mundepi, Srinivas Abhilash, Stuti Shekhar

Agenda

01/
Business
Objective

02/
Model Approach

03/
Data

- iShares Bond
- Natural Disaster Feature Extraction

04/
Model

- Model Results
- Model Evaluation
- Feature Importance

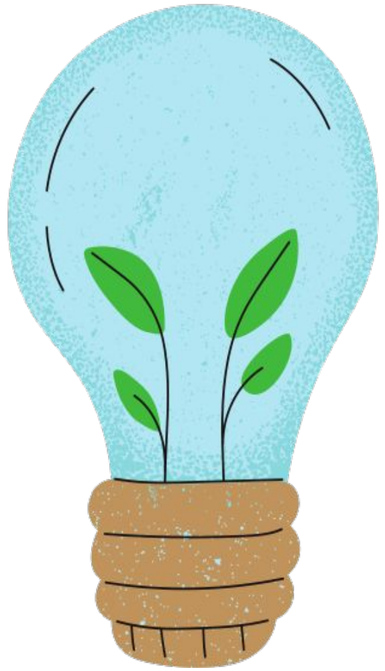
05/
Investment
Advisory Tool

- Scalability

06/
Benefits

01/ BUSINESS OBJECTIVE

To predict bond performance by analyzing environmental risks and aligning investment strategies. This involves identifying how environmental factors such as weather and natural disaster data can impact green bonds and making informed decisions to enhance returns.



02/ MODEL APPROACH

FEATURE ENGINEERING

Features extracted:

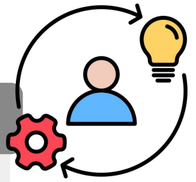
- Natural disaster features
- Cloud Coverage Rate

Image processing using template matching, grayscale conversion and thresholding

DATA PREPROCESSING

Handling data types, removing outliers, noise

MODELLING



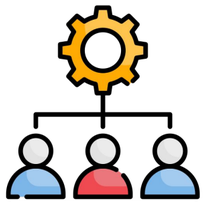
Training predictive models

- Logistic Regression
- Decision Tree
- Random Forest
- Gradient Boost
- Neural Network

Environmental and financial data sources

- BGRN bond data
- Visual Crossing API
- Image data from NASA

DATA SOURCE



EVALUATION

Hyper parameter tuning & Model selection using OOS R-square value

03/ DATA

Weather data (**22** variables)

temperature	dew	humidity	precipitation	windspeed	...

Bond

open	high	low	close	volume

Natural Disaster (**8** variables)

dust_haze	manmade	volcanoes	severe storm	wildfires	...

iShares BGRN

DEFINITION

The iShares Global Green Bond ETF (BGRN) is an exchange-traded fund that seeks to track the Bloomberg MSCI Global Green Bond Select Index.

PURPOSE

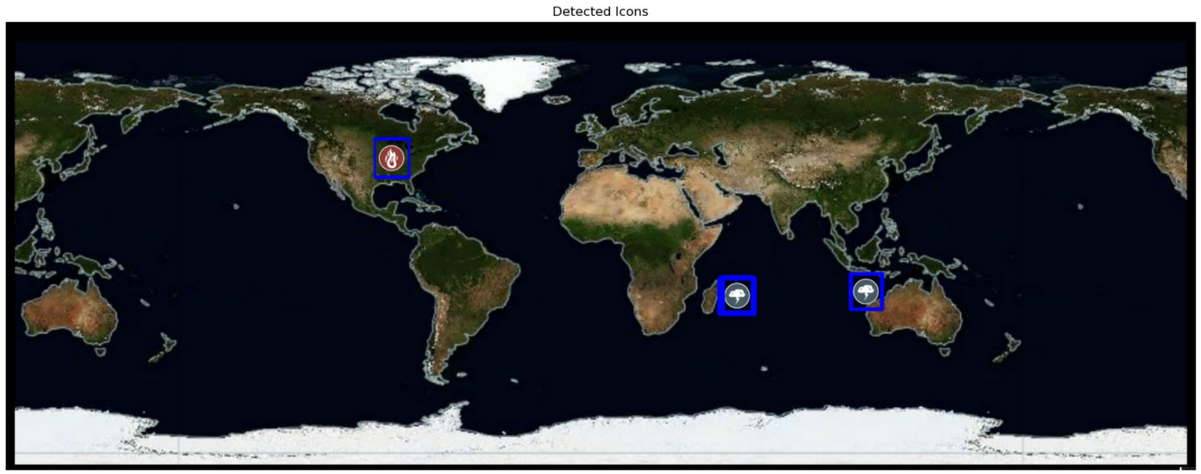
Invests in green bonds—bonds specifically issued to fund projects with positive environmental and climate benefits, such as renewable energy, energy efficiency, and clean transportation.

BENEFITS

- Sustainability
- Diversification
- Liquidity
- Transparency

Natural Disaster Feature Extraction

using template matching



TEMPLATES

Manmade

Dust_haze

Severe Storms

Water Color

Volcanoes

Snow

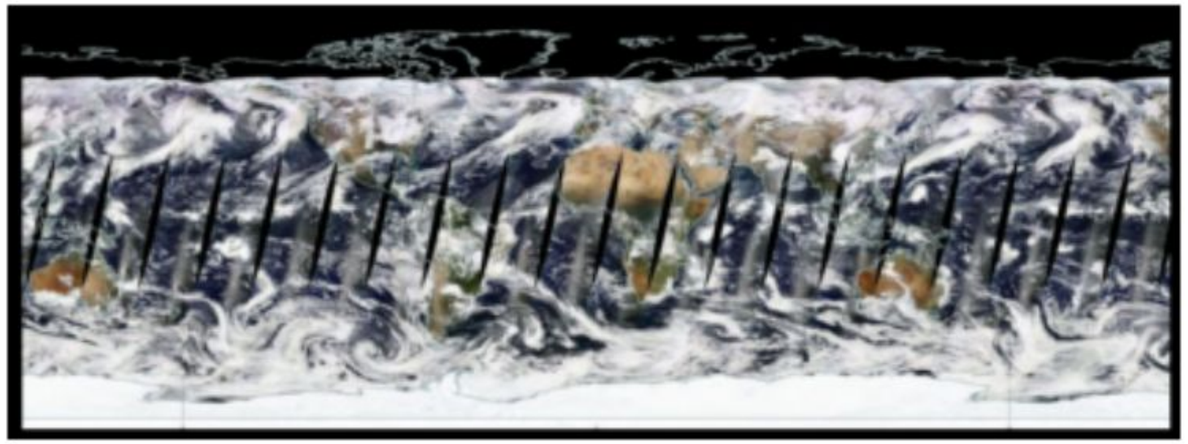
Wildfires

Sea and Lake Ice

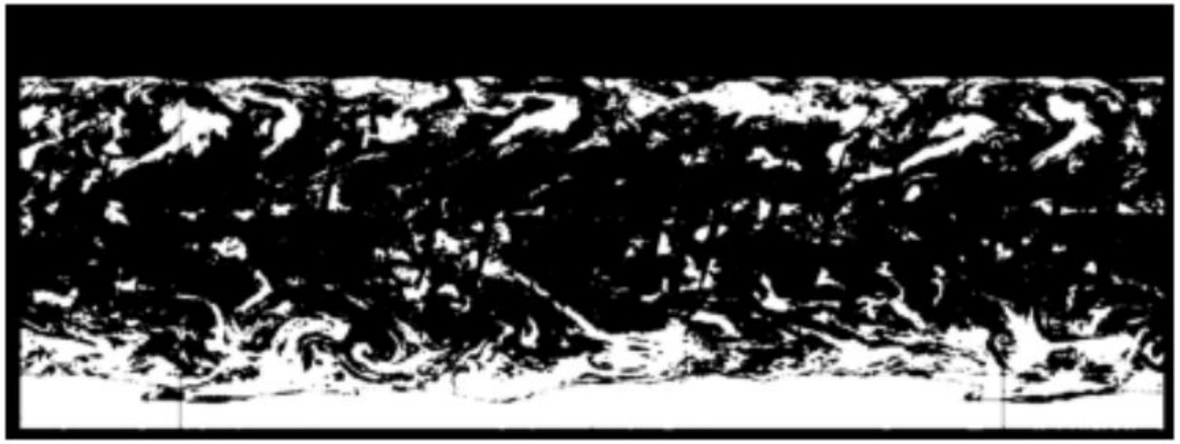
Cloud Coverage Estimation

Using Grayscale Conversion and Binary Thresholding

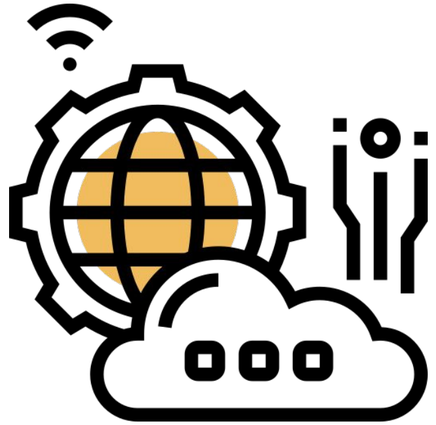
Input Image



Processed Image



Cloud cover: 26.51%



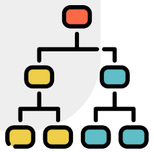
04/ MODEL

LOGISTIC REGRESSION

Logistic Regression is a statistical model used for binary classification that predicts the probability of an outcome based on input features.

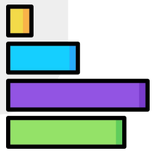
DECISION TREE

Decision Tree is a non-parametric supervised learning method used for classification and regression.



RANDOM FOREST

Random Forest constructs multiple decision trees during training and merges their outputs to improve predictive accuracy and control over-fitting.

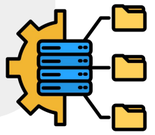


GRADIENT BOOSTING

Gradient Boosting builds an ensemble of weak prediction models, to improve accuracy by iteratively minimizing a loss function.

NEURAL NETWORK

Neural Network is inspired by the human brain, consisting of interconnected nodes that process data and learn patterns through multiple layers.

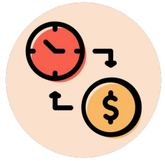


05/ MODEL RESULTS

DEPENDENT VARIABLE

Bond price for the following day

Models built using weather data
All the models had poor performance with negative R-squared values

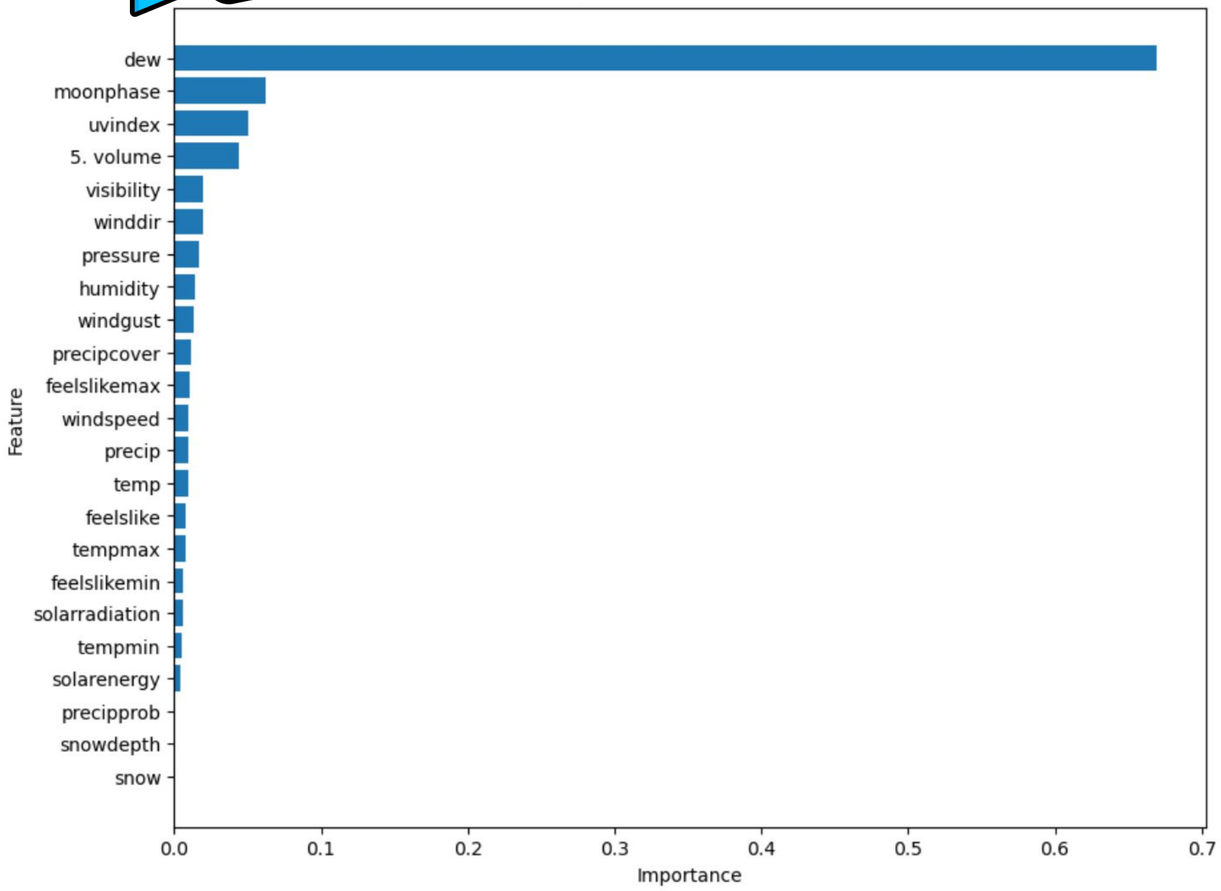


Models with enriched datasets (including natural disaster and cloud cover % features)
Enhanced model performance - Random Forest and XGBoost

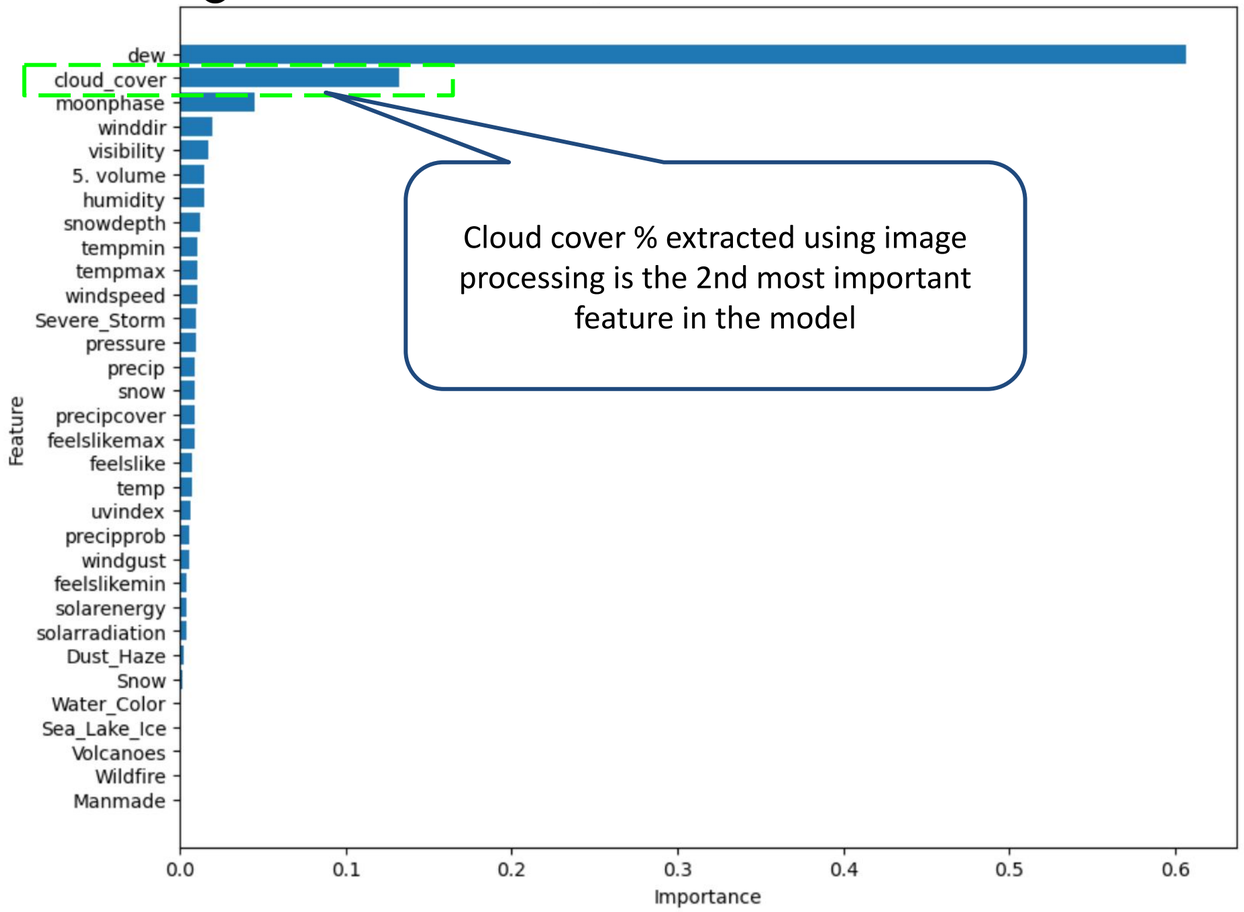
Technique	MSE	R-squared
Random Forest	0.005	45%
Gradient Boosting	0.007	19%

FEATURE IMPORTANCE

Feature Importance using weather data



Feature Importance using enriched data

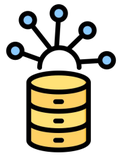


Cloud cover % extracted using image processing is the 2nd most important feature in the model

06/ SCALABILITY

DATA ENRICHMENT

Adding additional records & features



Historical Data

Extend the dataset by including more historical bond prices and related environmental variables

Temporal Expansion

Extend the dataset by including more granular time intervals (e.g., hourly or minute-level data instead of daily)

Spatial Data

If applicable, incorporate data from different regions or countries to capture geographical variations. This could lead up to **organisational issues** in managing the diverse regulatory and reporting standards .

Feature Engineering

- Include data on GDP, inflation rates, employment rates, and other macroeconomic indicators
- Include sentiment analysis from financial news, social media, or analyst reports

07/ GREEN BOND INVESTMENT ADVISORY TOOL

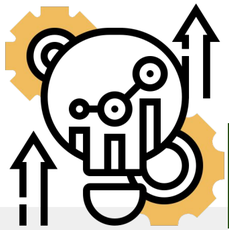
TARGET AUDIENCE

Retail Investors:

Individuals looking to invest in green bonds and other sustainable investment options.

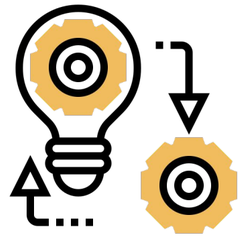
Financial Advisors:

Professionals seeking advanced tools to advise their clients on green investments.



KEY FEATURES

1. **Real-Time Bond Price Predictions:** Use ML & DL models to predict the price of iShares BGRN based on environmental data and natural disaster forecasts.
2. **Risk Assessment Dashboard:** Provides a comprehensive view of environmental risks associated with specific bonds, including the likelihood and potential impact of natural disasters.
3. **Portfolio Optimization & Personalized Investment Recommendations:** Use Advanced algorithms optimize bond portfolios as well as suggest investment strategies tailored to the user's risk tolerance, investment goals, and environmental preferences.
4. **Alerts and Notifications:** Sends real-time alerts about significant environmental events or changes in bond prices, and notifications about investment opportunities or risks.



08/ BENEFITS

- A **cutting-edge advisory tool** can differentiate our services from competitors, attracting clients interested in sustainable investing and retaining them through superior service.

- By providing **value-added services** such as personalized recommendations and portfolio optimization, we can charge premium fees or subscription models, enhancing revenue.

Revenue Estimation

Retail Investor Revenue

- **Total Retail Investors (Global):** 100 million (conservative estimate)
- **Adoption Rate:** 1%
- **Subscribers:** 1 million
- **Annual Revenue:** 1 million subscribers * $100/month * 12 months = $1.2 billion

Financial Advisor Revenue

- **Total Financial Advisors (Global):** 1 million (conservative estimate)
- **Adoption Rate:** 2%
- **Subscribers:** 20,000
- **Annual Revenue:** 20,000 subscribers * $500/month * 12 months = $120 million

Total Potential Revenue:

- **Retail Investors + Financial Advisors:** $1.2 billion + $120 million = **$1.32 billion annually**

THANK YOU !

06/ Benefits

Enhanced Financial Outcomes: Investors leverage predictive analytics to improve financial returns while aligning with environmental sustainability goals.

Promotion of Environmental Goals: The strategic use of green bonds demonstrates a commitment to environmental stewardship, influencing broader market trends towards sustainability.

Catalyst for Investment Shift: Encourages a shift in investor behavior, promoting greater consideration of environmental impacts in investment decisions and advancing the field of sustainable finance.

DECISION TREE

RMSE: 0.0648

The Root Mean Squared Error (RMSE) of the Decision Tree model is 0.0648, indicating a good fit with the data.

R^2: 0.8613

The R-squared (R^2) value of 0.8613 suggests that the Decision Tree model can explain 86.13% of the variance in the target variable.

Slightly higher RMSE compared to Linear Regression

While the RMSE of the Decision Tree model is slightly higher than the Linear Regression model, it still performs well in predicting the BGRN prices.

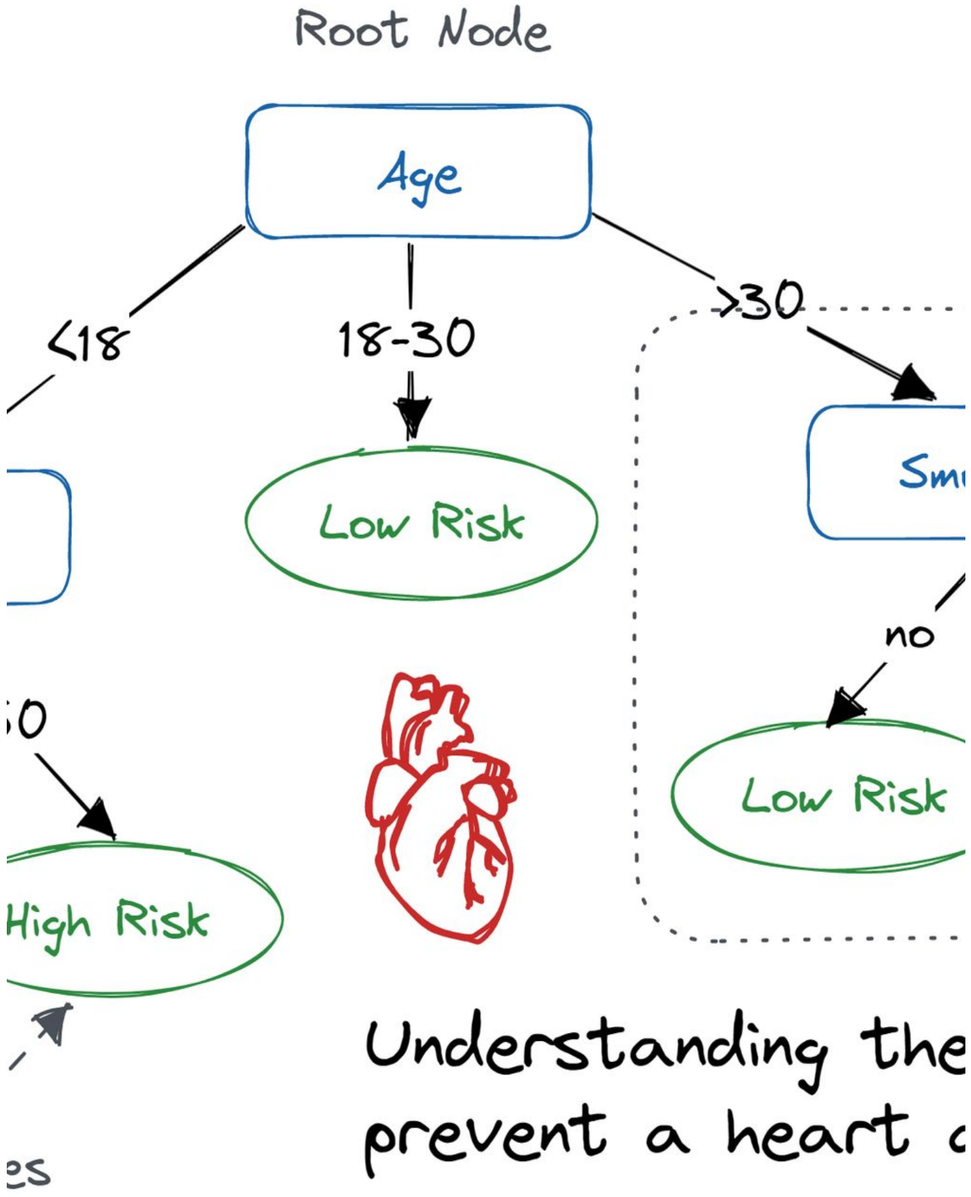
The Decision Tree model demonstrates strong performance in predicting BGRN prices, with a high R-squared value and a relatively low RMSE, making it a viable option for price forecasting.

Decision Tree

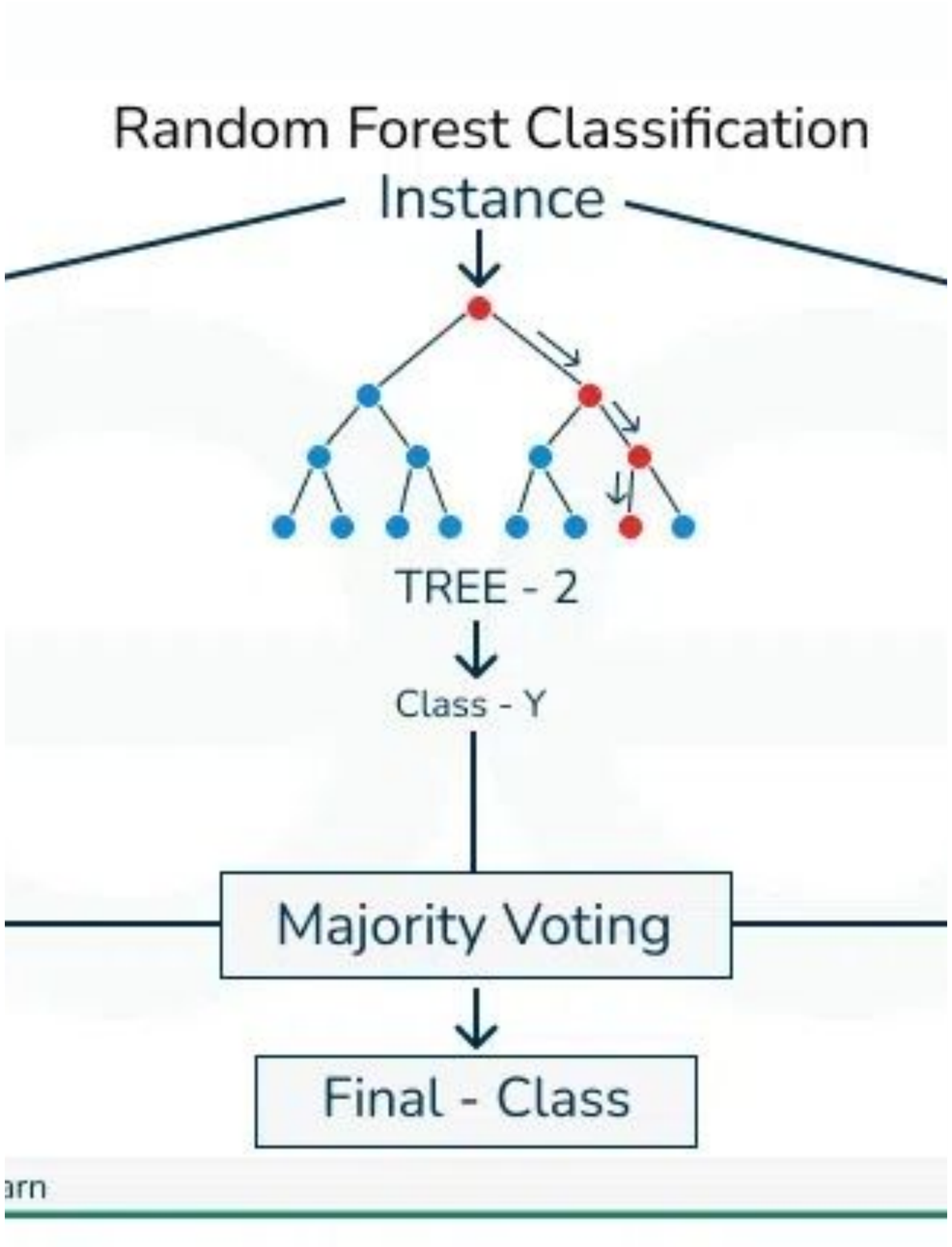
RMSE: 0.0648

R^2: 0.8613

Slightly higher RMSE compared to Linear Regression



Random Forest



Gradient Boosting

RMSE: 0.0588

The Root Mean Squared Error (RMSE) of the Gradient Boosting Model is 0.0588, indicating a relatively small error in its predictions.

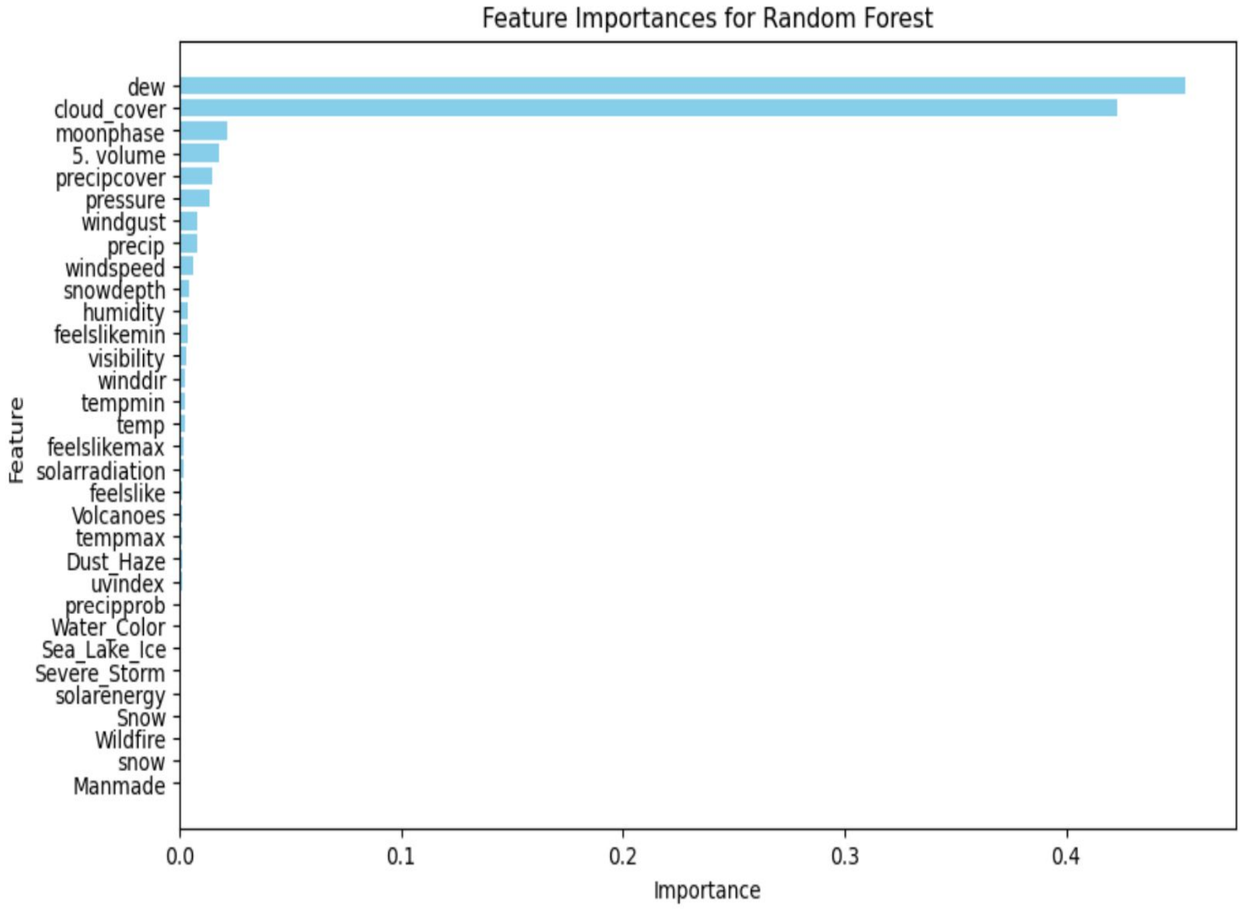
R^2: 0.8856

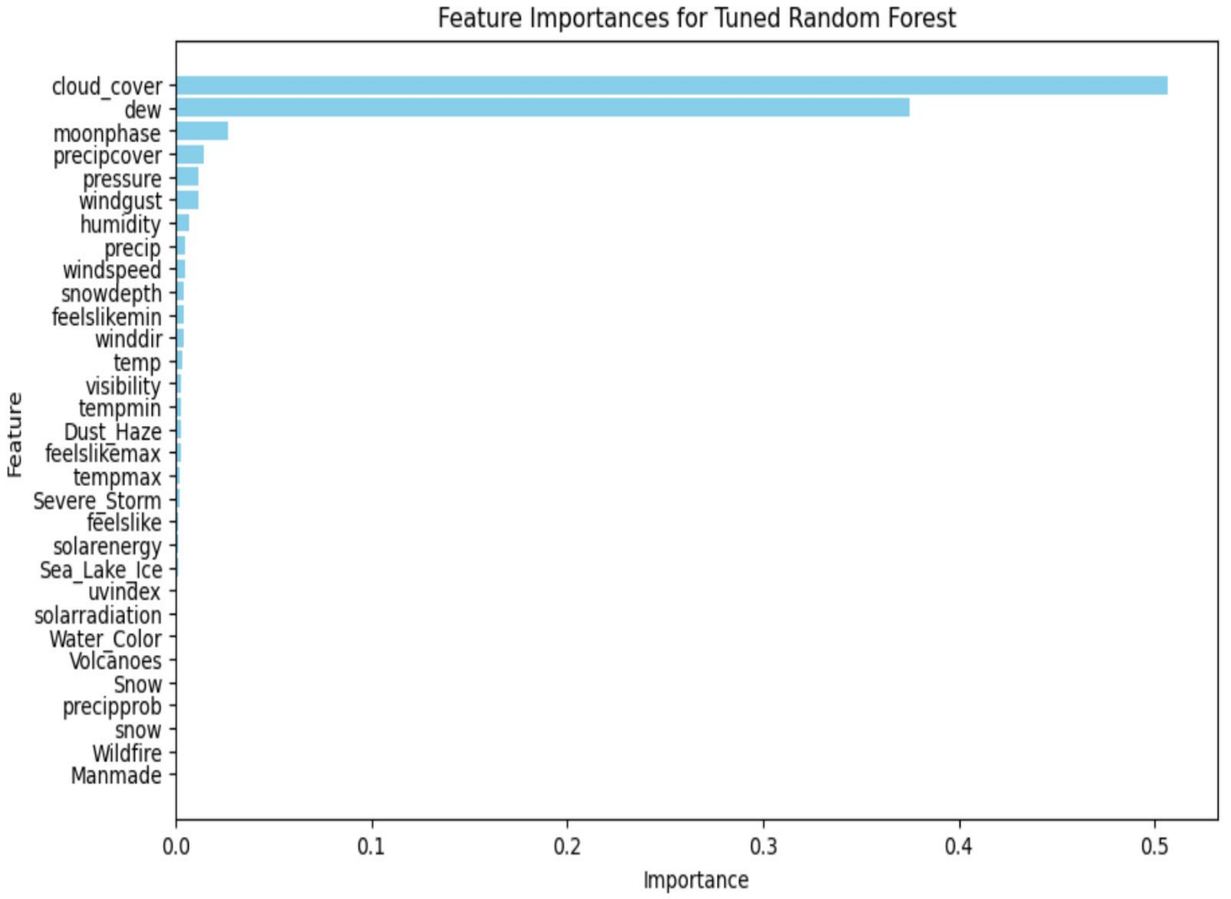
The R-squared value of 0.8856 suggests that the Gradient Boosting Model explains a large proportion of the variance in the target variable.

Performs similarly to Random Forest

The Gradient Boosting Model performs comparably to the Random Forest model, with a slightly higher RMSE but still very robust.

The Gradient Boosting Model demonstrates strong performance metrics, making it a viable choice for predictive modeling tasks.

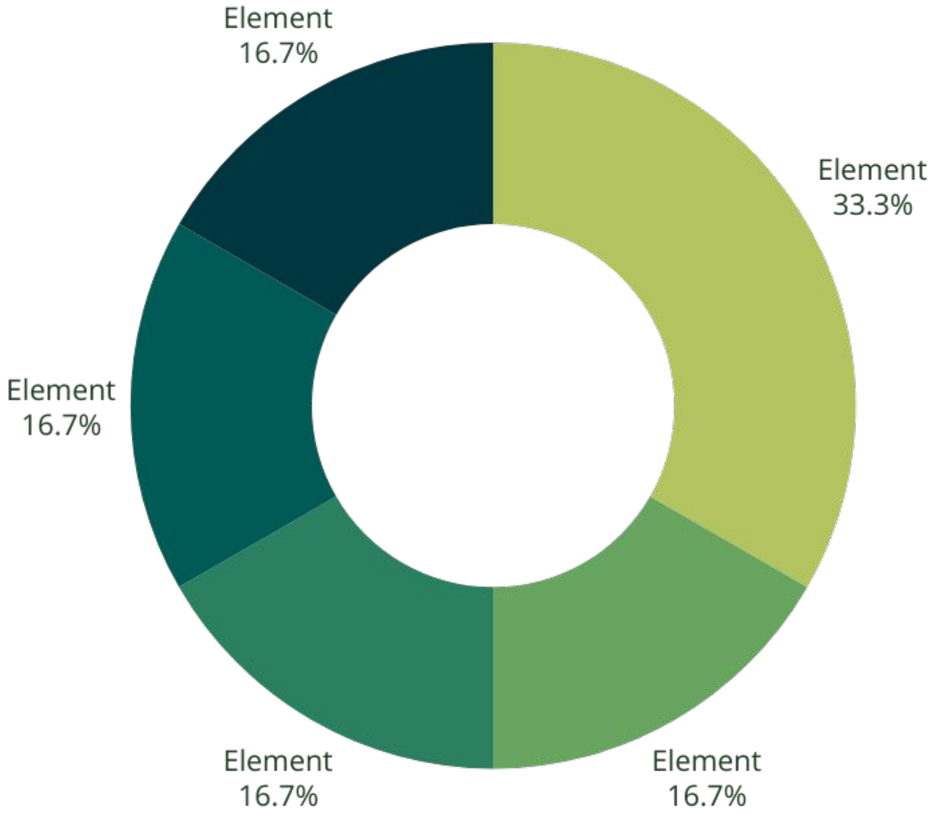

Feature Importances for Random Forest
dew
cloud_cover
moonphase
5. volume
precipcover
pressure
windgust
precip
windspeed
snowdepth
humidity
feelslikemin
visibility
winddir
tempmin
temp
feelslikemax
solarradiation
feelslike
Volcanoes
tempmax
Dust_Haze
uvindex
precipprob
Water_Color
Sea_Lake_Ice
Severe_Storm
solarenergy
Snow
Wildfire
snow
Manmade
Feature
0.0
0.1
0.2
0.3
0.4
Importance


Feature Importances for Tuned Random Forest
cloud_cover
dew
moonphase
precipcover
pressure
windgust
humidity
precip
windspeed
snowdepth
feelslikemin
winddir
temp
visibility
tempmin
Dust_Haze
feelslikemax
tempmax
Severe_Storm
feelslike
solarenergy
Sea_Lake_Ice
uvindex
solarradiation
Water_Color
Volcanoes
Snow
precipprob
snow
Wildfire
Manmade
Feature
0.0
0.1
0.2
0.3
0.4
0.5
Importance

01/

Be sustainable

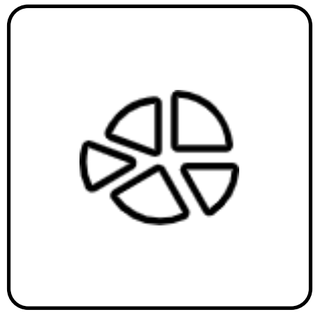
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.



Linear Regression

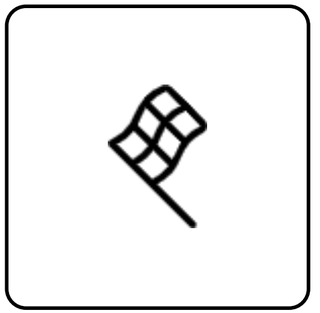
RMSE: 0.0625

The Root Mean Squared Error (RMSE) of 0.0625 indicates a relatively low error rate in the model's predictions, suggesting a good fit to the data.



R^2: 0.8708

The R-squared value of 0.8708 shows that the model explains 87.08% of the variance in the target variable, indicating a strong predictive power.



Solid Baseline

The combination of a low RMSE and high R^2 provides a solid baseline model for the BGRN price prediction task, offering a reliable starting point for further model refinement.

The linear regression model evaluated exhibits a strong performance, with a low RMSE and high R^2, making it a suitable baseline for the BGRN price prediction.