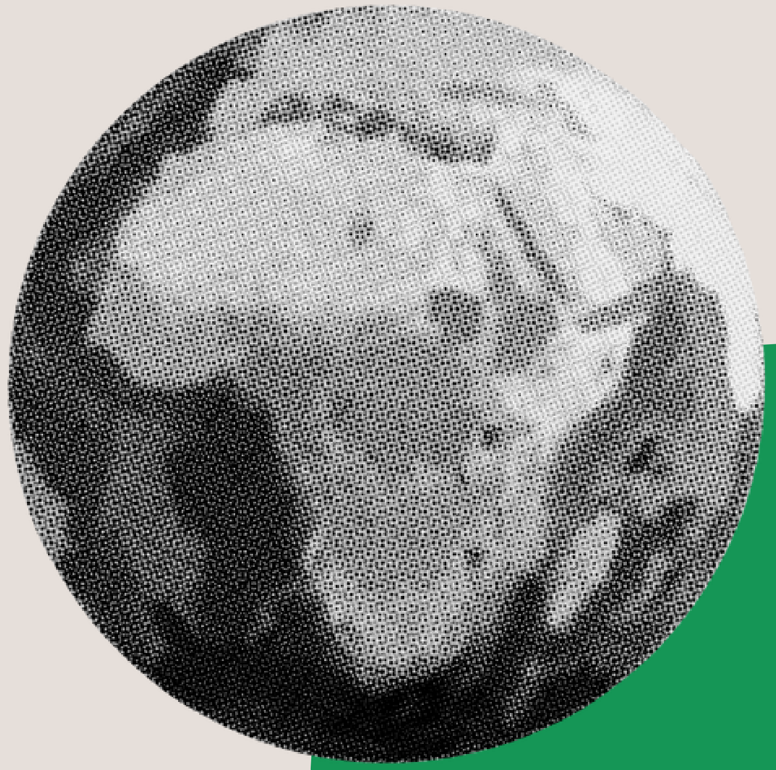


SUPERVISED LEARNING -CLASSIFICATION

Presented by: Saifuddin Malik
Yogendran
Osborm

Weather Type for Resource Management and Predictive Maintenance

Table Of Content

Weather Type for Resource Management and Predictive Maintenance

Introduction

WHAT	Weather directly impacts farming activities, crop health, and resource utilization
WHY	Unpredictable weather patterns can lead to inefficiencies, increased costs, and potential crop losses
HOW	Utilizing advanced machine learning models to classify and predict weather conditions

What’s the Problem Definition?

Accurate weather forecasts are essential for farmers to optimize irrigation, maintain equipment, and manage resources effectively, helping to reduce costs and improve overall efficiency in the face of unpredictable weather.

Overview:

What’s the objective/Goal?

The goal is to help farmers optimize resource management and equipment maintenance through accurate weather predictions, enabling data-driven decisions that boost efficiency, reduce costs, and enhance preparedness for adverse weather.

What’s the Problem Statement?

To develop a reliable weather classification model, we can use weather-related features (e.g., temperature, humidity, wind speed) to predict conditions like Rainy, Sunny, Cloudy, and Snowy. This model will support farmers in making informed decisions on resource allocation and equipment maintenance, improving efficiency and reducing costs.

Data Collection

Source Of Data

Description of Data

Weather Type Classification

Forecast with Precision: Simulated Data for Predicting Weather Types

k kaggle.com

Description of Data

Feature	Type	Description
Temperature	Numeric	The temperature in degrees Celsius, ranging from extreme cold to extreme heat.
Humidity	Numeric	The humidity percentage, including values above 100% to introduce outliers.
Wind Speed	Numeric	The wind speed in kilometers per hour, with a range including unrealistically high values.
Precipitation (%)	Numeric	The precipitation percentage, including outlier values.
Cloud Cover	Categorical	The cloud cover description.
Atmospheric Pressure	Numeric	The atmospheric pressure in hPa, covering a wide range.
UV Index	Numeric	The UV index, indicating the strength of ultraviolet radiation.
Season	Categorical	The season during which the data was recorded.
Visibility (km)	Numeric	The visibility in kilometers, including very low or very high values.
Location	Categorical	The type of location where the data was recorded.
Weather Type	Categorical	The target variable for classification, indicating the weather type (Rainy, Sunny, Cloudy, Snowy).

Data Pre-Processing

Data Wrangling

Encoding

Data Augmentation:

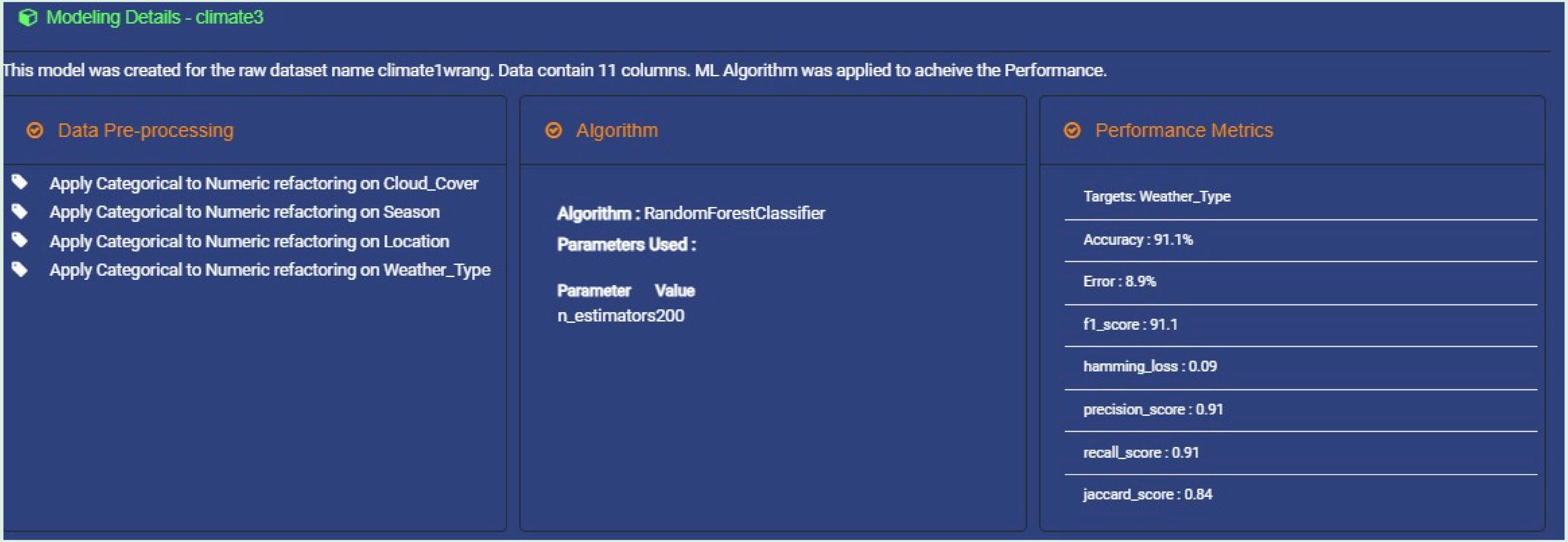
- **Encoding:** Converts categorical features into numerical values to make them compatible with machine learning algorithms.
- **Data Augmentation:** Synthetically generates new data samples to increase dataset diversity and balance. (Test & Train)

EDA (Exploratory Data Analysis)

Data Visualizations **Correlation Analysis**

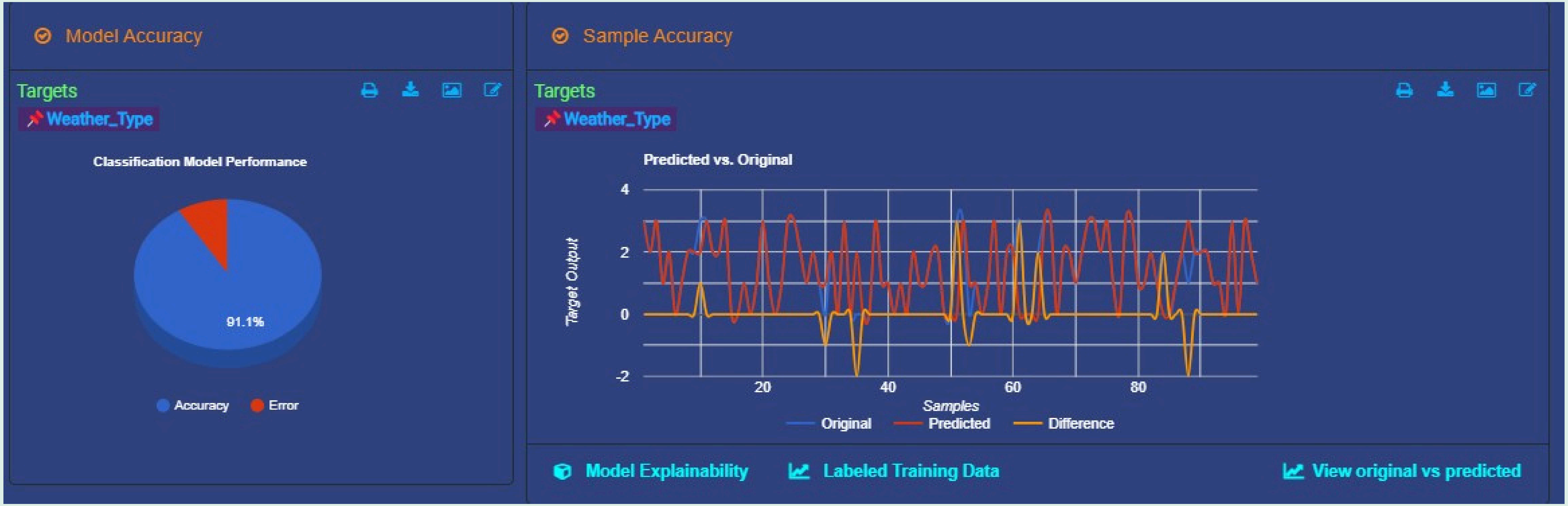
- Data Visualizations: Uses charts to display relationships, trends, and outliers visually.
- Correlation Analysis: Identifies strong relationships between variables to guide feature selection.

Modeling Details: Methodlogy



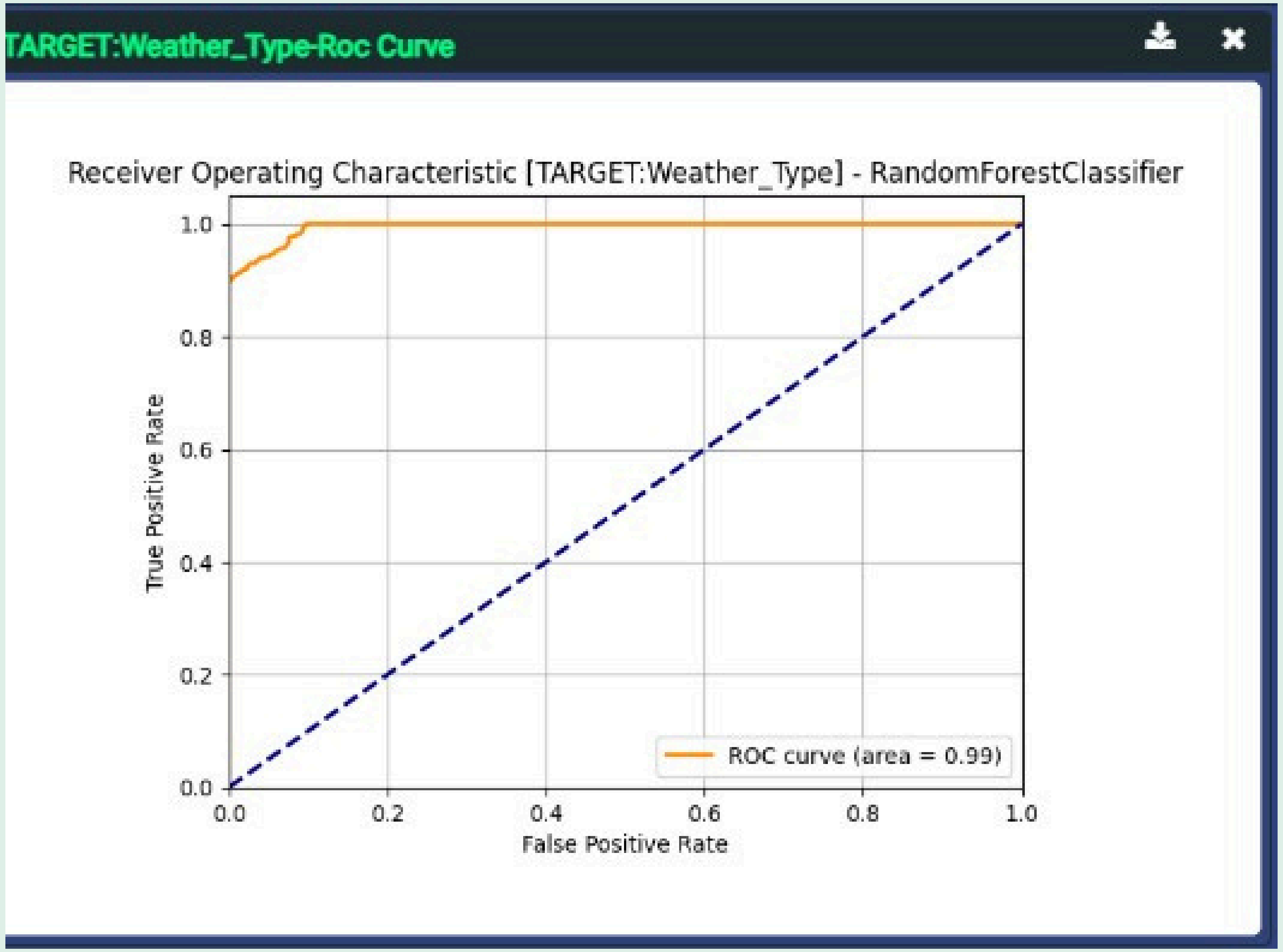
- This statistics summary shows the results of a Random Forest Classifier applied to our dataset. The categorical features such as Cloud Cover, Season, Location, and Weather Type were converted to numeric values. The model achieved a high accuracy of 91.1% with an F1-score of 91.1, precision score of 0.91, and recall score of 0.91, indicating strong overall performance in predicting the target, `Weather_Type`

Model Accuracy & Sample Accuracy :



- This data visualization highlights the model's accuracy in predicting the `Weather_Type` target. The pie chart on the left shows that the model achieved 91.1% accuracy, with a small error portion. The line graph on the right compares the predicted values against the original values, showing a close match with some differences, indicating the model is effective.

Roc Curve



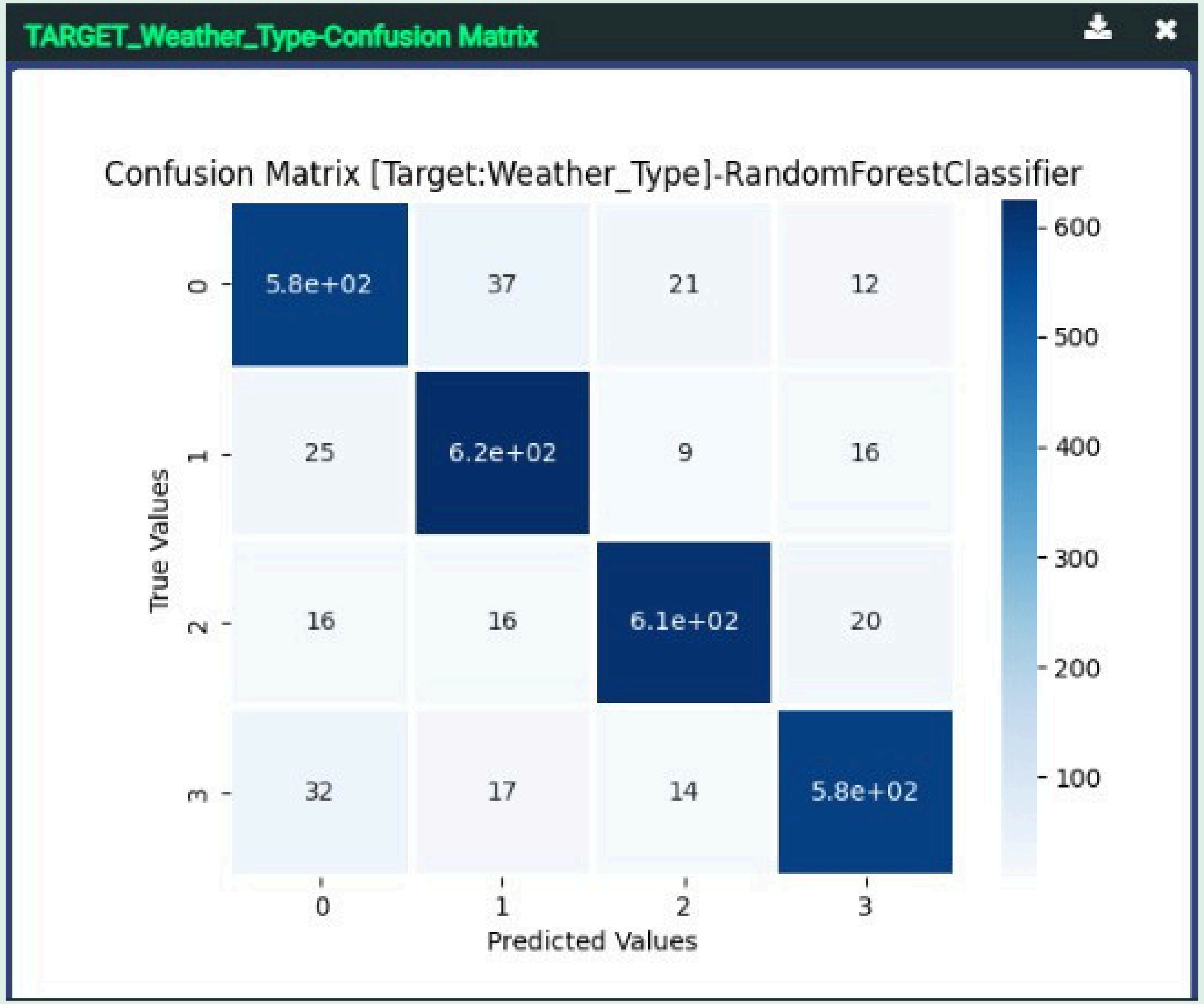
1. The ROC curve evaluates the performance of a RandomForestClassifier predicting "Weather_Type".

2. The orange curve shows a high true positive rate and a low false positive rate, indicating strong model performance.

3. The diagonal dotted line represents random estimation that the model's curve is much better, close to the top left corner.

4. The Area Under the Curve (AUC) is 0.99, signifying near perfect classification accuracy.

Confusion Matrx:



1. Rows represent actual "Weather_Type" values, and columns represent predicted values.

2. The diagonal cells contain the counts where the model correctly predicted the weather type, with high values (e.g., 580, 620, 610) indicating good accuracy.

3. These off diagonal values cells show where the model made errors, predicting a different weather type than the actual one. Lower values here indicate fewer mistakes.

4. Darker cells represent higher counts, with most of the diagonal cells being dark, reinforcing the model's accuracy in predictions.

Results & Conclusion

Classification	Accuracy (%)	AOC	Error (%)
Random Forest Classifier - 100	90.83	0.99	9.17
Decision Tree Classifier	90.04	0.94	9.96
Support Vector Machine	79.13	0.99	20.87
Random Forest Classifier - 200	91.1	0..99	8.9

Recommendations:

Integration into Farm Management Systems:
The weather classification model can be integrated into existing farm management software used by agricultural organizations. This allows for real-time weather predictions that can be used for decision-making, such as scheduling irrigation, planting, and harvesting.

Mobile and Web Applications:
Develop mobile and web applications for farmers to access weather predictions easily. This would provide them with timely and actionable insights, enhancing their ability to respond to changing weather conditions.

Results

The Random Forest Classifier achieved 91.1% accuracy with strong metrics across F1-score, precision, recall, and an AUC of 0.99, showing effective weather classification.

Conclusion

The model can enhance agricultural efficiency and reduce costs, with potential improvements in using real-world data and advanced feature engineering.

Lesson Learned

The inclusion of synthetic outliers provided valuable insights into the model's robustness. Improvements could include using real-world data, enhancing feature engineering, and applying advanced outlier detection techniques to further refine the model.

THANK YOU