



Project Report: SmartML Explorer - Automated

Machine Learning with PyCaret

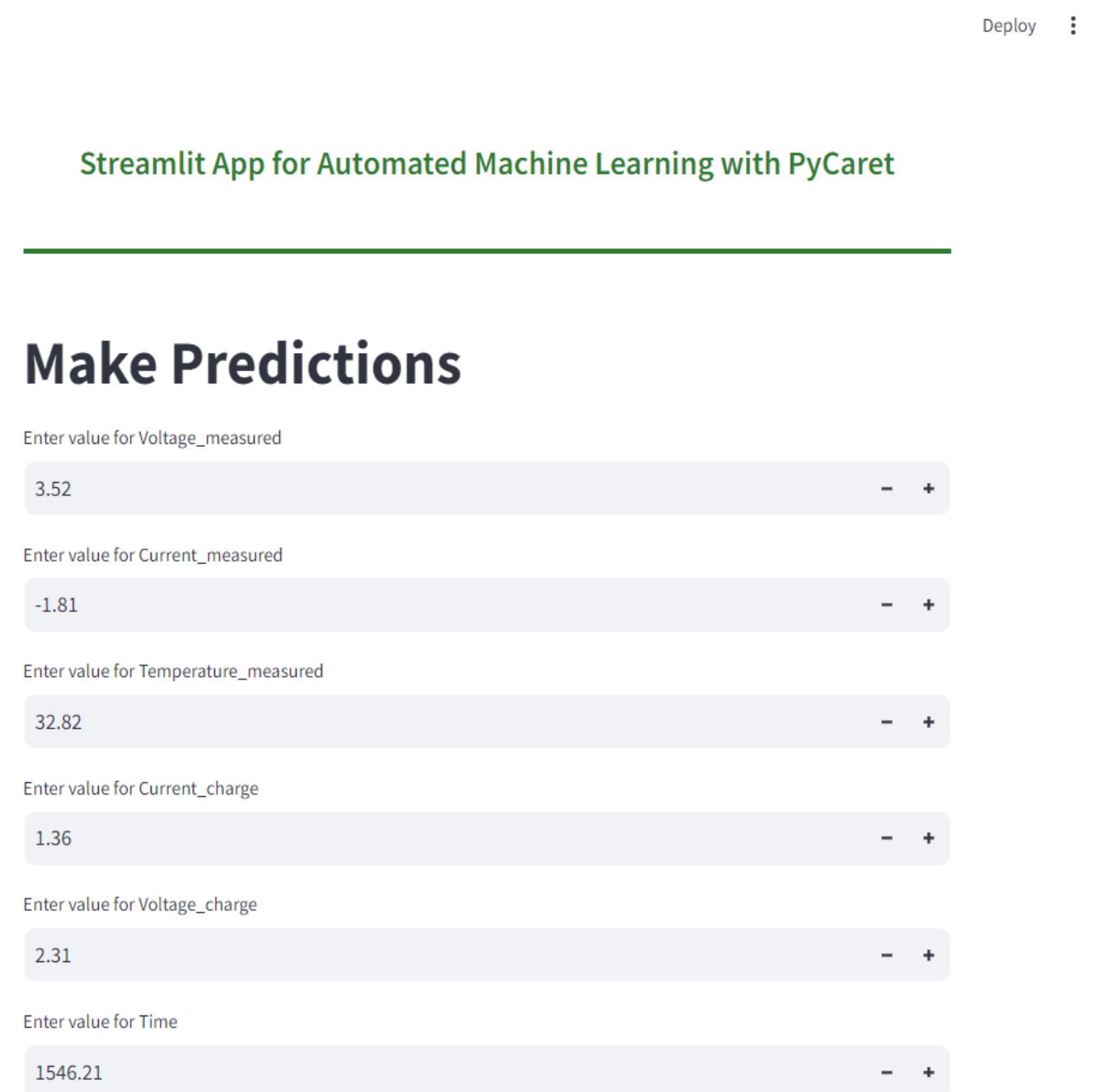
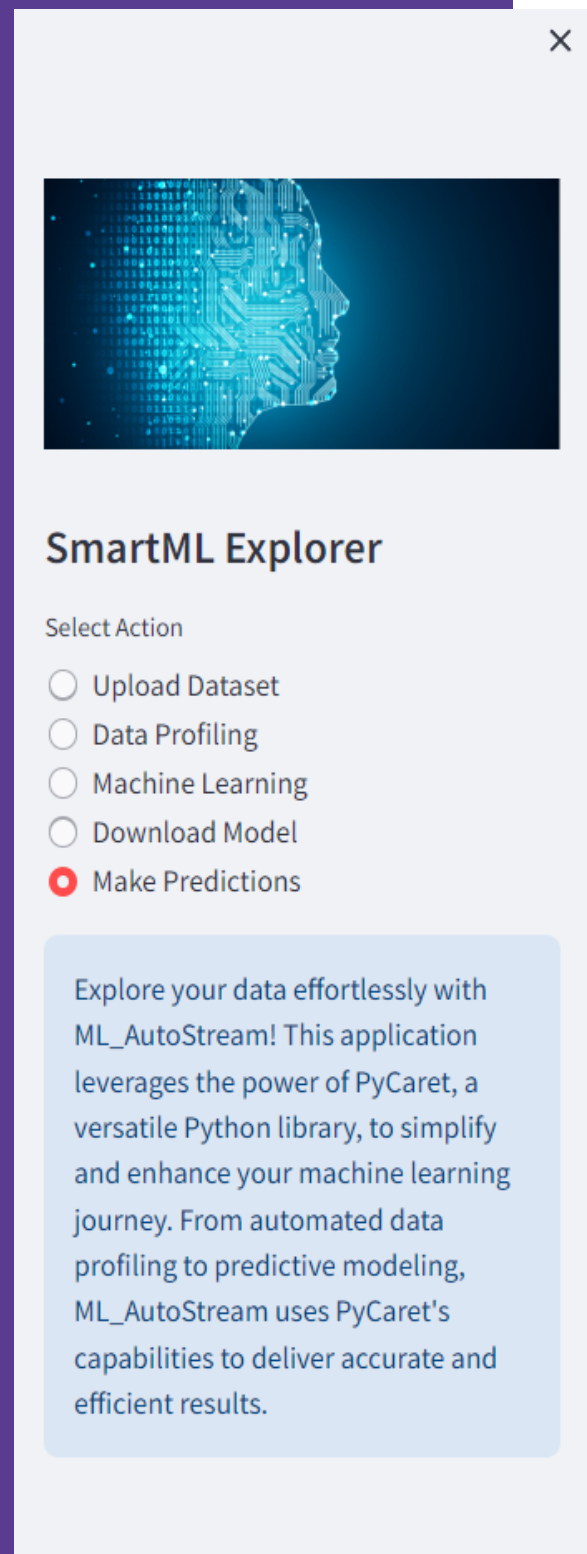
OVERVIEW



SmartML Explorer is a Streamlit application designed for automated machine learning using the PyCaret library. This report provides a comprehensive overview of the application's functionalities and highlights its key features.



The project aims to simplify the machine learning journey by leveraging PyCaret's capabilities within a Streamlit app. Users can easily upload datasets, perform automated data profiling, explore machine learning models, and make predictions, all through an intuitive interface.



PROGRAM & TOOLS

Programs:

Python, Visual Studio Code, Streamlit

Libraries:

Pandas: Data manipulation library for reading, analyzing, and manipulating datasets.

YData Profiling: Automated exploratory data analysis library for generating detailed reports on dataset characteristics.

Streamlit YData Profiling: Integrates YData Profiling with Streamlit for displaying automated profiling reports.

PyCaret: Open-source machine learning library for simplified ML workflows.

Matplotlib: Python 2D plotting library for creating visualizations.

Machine Learning Specific:

setup, compare_models, pull, save_model, load_model, predict_model (PyCaret): PyCaret functions for ML experiments and predictions.


Visualization:


plt (Matplotlib): Matplotlib's pyplot module for creating visualizations.

Code Structure

Data Exploration and Profiling:

Users can gain insights into their datasets through automated profiling and visualizations






SmartML Explorer

Select Action

- ☒ Upload Dataset
- ☐ Data Profiling
- ☐ Machine Learning
- ☐ Download Model
- ☐ Make Predictions


Explore your data effortlessly with ML_AutoStream! This application leverages the power of PyCaret, a versatile Python library, to simplify and enhance your machine learning journey. From automated data profiling to predictive modeling, ML_AutoStream uses PyCaret's capabilities to deliver accurate and efficient results.



Streamlit App for Automated Machine Learning with PyCaret


Upload


Upload your dataset here



Drag and drop file here
Limit 200MB per file

Browse files

 B0005_discharge.csv 5.9MB




	Voltage_measured	Current_measured	Temperature_measured	Current_charge	Voltage_charge	Tim
0	4.1915	-0.0049	24.33	-0.0006	0	
1	4.1907	-0.0015	24.326	-0.0006	4.206	16
2	3.9749	-2.0125	24.3891	-1.9982	3.062	35
3	3.9517	-2.014	24.5448	-1.9982	3.03	53
4	3.9344	-2.0111	24.7314	-1.9982	3.011	71
5	3.9201	-2.013	24.9098	-1.9982	2.991	90
6	3.9079	-2.0144	25.1059	-1.9982	2.977	108
7	3.897	-2.0116	25.317	-1.9982	2.967	126
8	3.8875	-2.018	25.5004	-1.9982	2.958	144

Code Structure

Data Exploration and Profiling:

Users can gain insights into their datasets through automated profiling and visualizations



SmartML Explorer

Select Action

- ☐ Upload Dataset
- ☒ Data Profiling
- ☐ Machine Learning
- ☐ Download Model
- ☐ Make Predictions

Explore your data effortlessly with ML_AutoStream! This application leverages the power of PyCaret, a versatile Python library, to simplify and enhance your machine learning journey. From automated data profiling to predictive modeling, ML_AutoStream uses PyCaret's capabilities to deliver accurate and efficient results.

Automated exploratory data analysis

Overview

Overview

Alerts **14**

Reproduction

Dataset statistics

Number of variables	12
Number of observations	50285
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	4.6 MiB
Average record size in memory	96.0 B

Variable types

Numeric	8
Categorical	4

Deploy

Data Exploration and Profiling

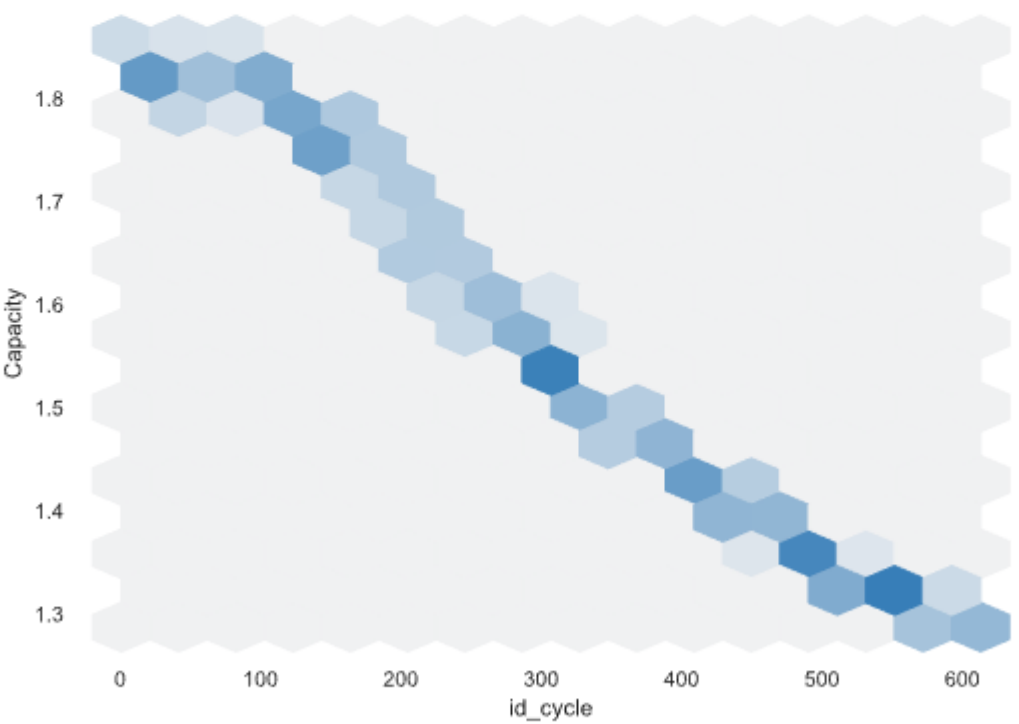
Interactions

Voltage_measured Current_measured Temperature_measured Current_charge

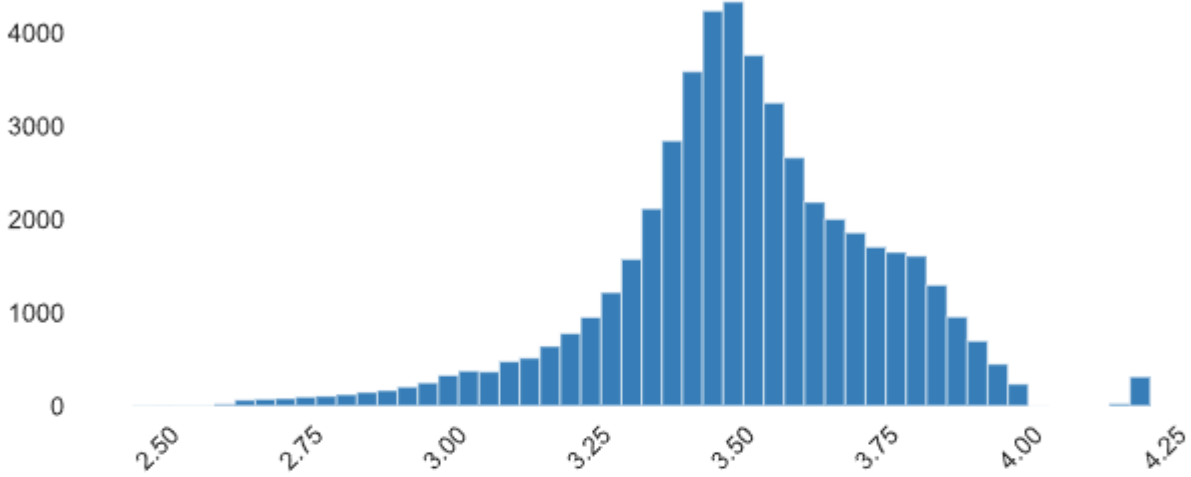
Voltage_charge Time Capacity id_cycle

id_cycle Voltage_measured Current_measured Temperature_measured

Current_charge Voltage_charge Time Capacity



stics Histogram Common values Extreme values



Histogram with fixed size bins (bins=50)

rent_measured

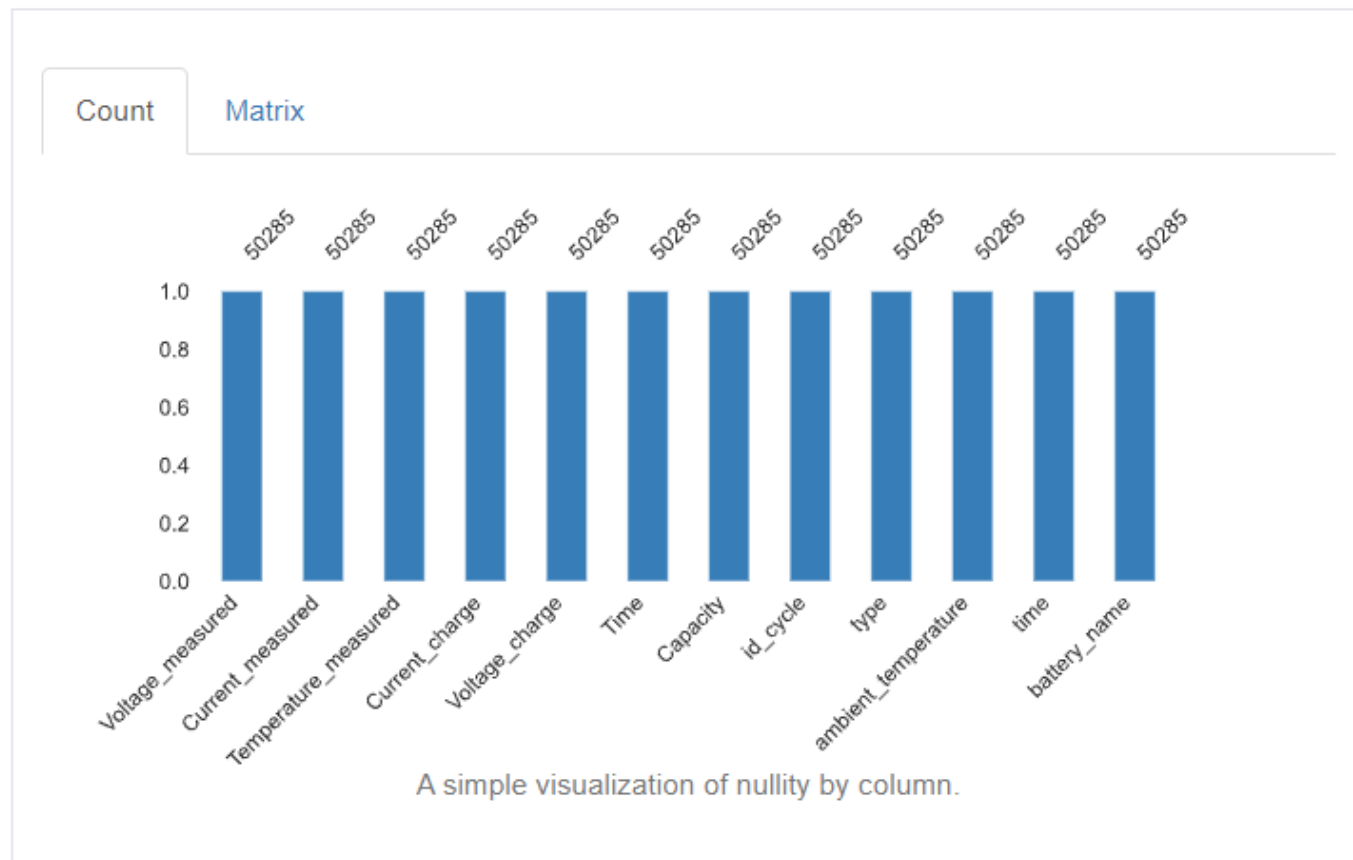
number (ℝ)

UE

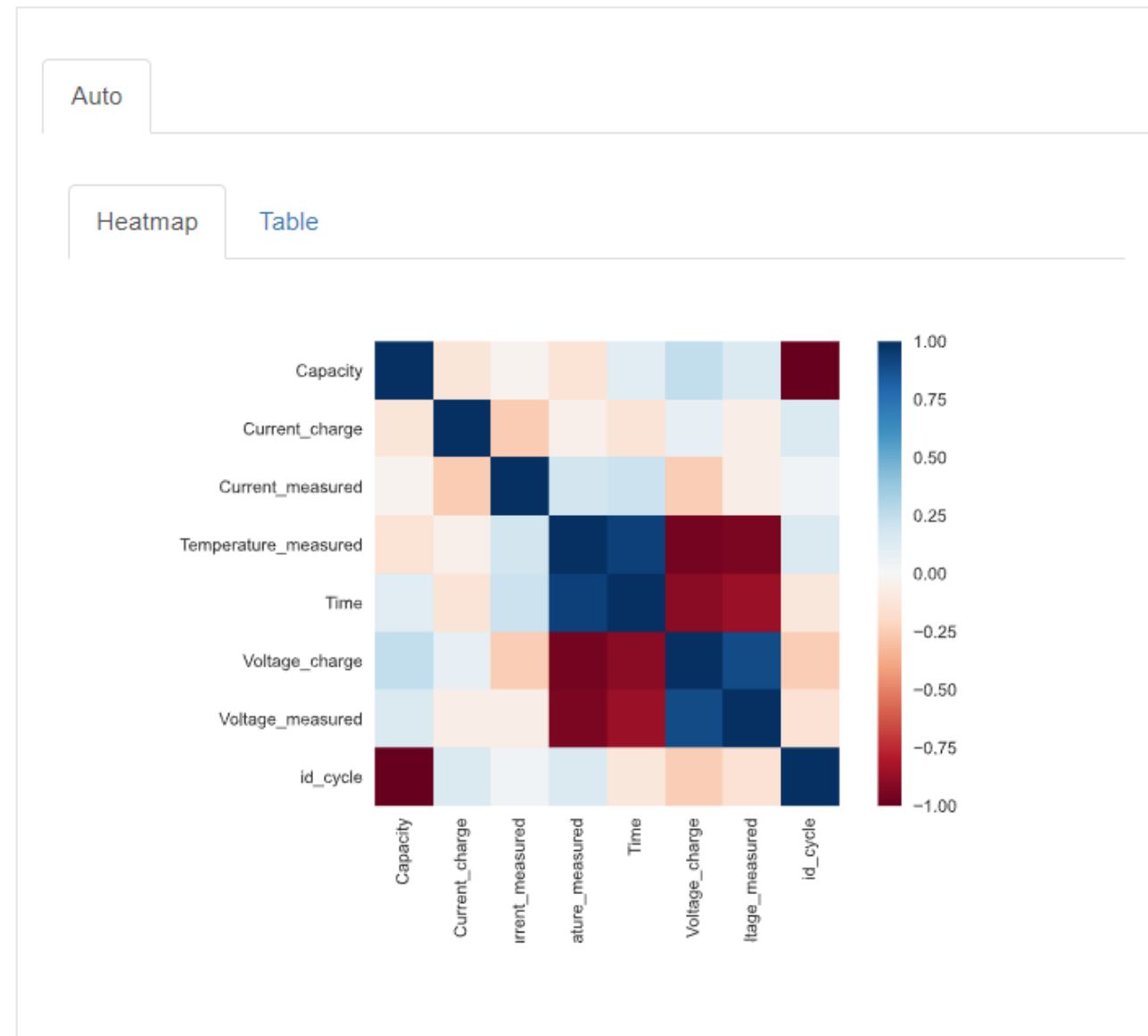
inct	50285
inct (%)	100.0%
sing	0
sing (%)	0.0%
rite	0
rite (%)	0.0%
in	-1.8060318

Data Exploration and Profiling

Missing values




Correlations



Machine Learning Exploration:

- Users select a target variable for machine learning exploration.



SmartML Explorer

Select Action

☐ Upload Dataset

☐ Data Profiling

☒ Machine Learning

☐ Download Model

☐ Make Predictions

Explore your data effortlessly with ML_AutoStream! This application leverages the power of PyCaret, a versatile Python library, to simplify and enhance your machine learning journey. From automated data profiling to predictive modeling, ML_AutoStream uses PyCaret's capabilities to deliver accurate and efficient results.

Code Structure

Streamlit App for Automated Machine Learning with PyCaret

Machine Learning - Exploration and Model Comparison

Select your target

Voltage_measured

train model

Machine Learning Experiment Settings:

	Description	Value
0	Session id	4860
1	Target	Voltage_measured
2	Target type	Regression
3	Original data shape	(50285, 12)
4	Transformed data shape	(50285, 12)
5	Transformed train set shape	(35199, 12)

Code Structure

Machine Learning Exploration:

- Experiment settings and best-performing models are displayed

Best Machine Learning Model:

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
br	Bayesian Ridge	0.0451	0.0051	0.0715	0.9053	0.017	0.0135	0.038
ridge	Ridge Regression	0.0451	0.0051	0.0715	0.9053	0.017	0.0135	0.037
lr	Linear Regression	0.0451	0.0051	0.0715	0.9053	0.017	0.0135	0.646
huber	Huber Regressor	0.044	0.0066	0.0814	0.8772	0.0196	0.0134	0.148
en	Elastic Net	0.0728	0.015	0.1225	0.7215	0.0287	0.0217	0.045
lasso	Lasso Regression	0.0732	0.015	0.1226	0.721	0.0288	0.0218	0.046
llar	Lasso Least Angle Regression	0.0732	0.015	0.1226	0.721	0.0288	0.0218	0.033
par	Passive Aggressive Regressor	0.0835	0.0168	0.1292	0.6896	0.0303	0.0249	0.035
omp	Orthogonal Matching Pursuit	0.0952	0.0179	0.1338	0.668	0.0311	0.0282	0.033
dummy	Dummy Regressor	0.1733	0.0539	0.2322	-0.0004	0.0523	0.0503	0.031

The model below is the best-performing model.

```
ExtraTreesRegressor
ExtraTreesRegressor(n_jobs=-1, random_state=4860)
```

Model Download



SmartML Explorer

Select Action

- ☐ Upload Dataset
- ☐ Data Profiling
- ☐ Machine Learning
- ☒ Download Model
- ☐ Make Predictions

Streamlit App for Automated Machine Learning with PyCaret

Download the model

download the model

Making Predictions



SmartML Explorer

Select Action

- ☐ Upload Dataset
- ☐ Data Profiling
- ☐ Machine Learning
- ☐ Download Model
- ☒ Make Predictions

Explore your data effortlessly with ML_AutoStream! This application leverages the power of PyCaret, a versatile Python library, to simplify and enhance your machine learning journey. From automated data profiling to predictive modeling, ML_AutoStream uses PyCaret's capabilities to deliver accurate and efficient results.

Streamlit App for Automated Machine Learning with PyCaret

Make Predictions

Enter value for Voltage_measured

3.52

- +

Enter value for Current_measured

-1.81

- +

Enter value for Temperature_measured

32.82

- +

Enter value for Current_charge

1.36

- +

Enter value for Voltage_charge

2.31

- +

Enter value for Time

1546.21

- +

Making Predictions:

Users input values for prediction, and the app displays the predicted outcomes

Making Predictions

Enter value for ambient_temperature

24.00

- +

Enter value for time

2008.00

- +

Select value for battery_name

B0005

▼

Select X-Axis and Y-Axis for plot:

Select a column for the x axis

id_cycle

▼

Select a column for the y axis

Capacity

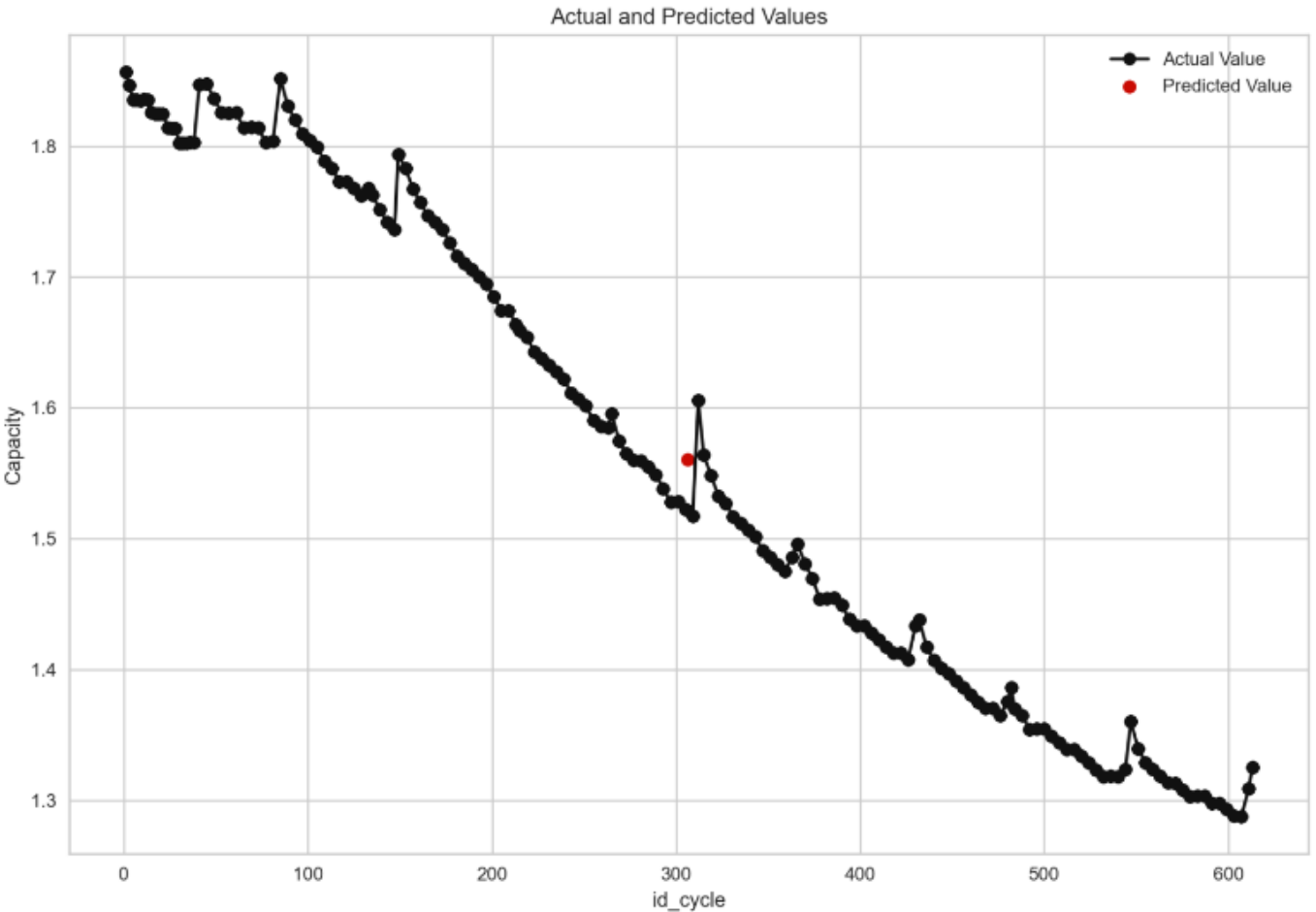
▼

Predict

	Voltage_measured	Current_measured	Temperature_measured	Current_charge	Voltage_charge	Tim
0	3.5153	-1.806	32.817	1.3627	2.3084	1,54

The prediction is:

	Capacity
0	1.5603



Making Predictions:

Users input values for prediction, and the app displays the predicted outcomes



THANK YOU



zosk62@gmail.com